ENCYCLOPEDIA OF

Physical Science and Technology

THIRD EDITION

Organic Chemistry

Table of Contents
(Subject Area: Organic Chemistry)

Acetylene

Robert J. Tedeschi
Tedeschi and Associates, Inc.

GLOSSARY

Acetylenic Pertaining to organic compounds containing a triple bond (—C≡C—) or acetylene group in the molecule.

Adjuvant Acetylenic diol used with pesticides to enhance activity, lower the rate of application, and increase safety.

Alkynol Primary, secondary, or tertiary acetylenic alcohol with the hydroxyl group generally adjacent or α to the triple bond.

Ammoxidation Acrylonitrile manufactured by oxidation of propylene in the presence of ammonia.

Aprotic Pertaining to highly polar solvents such as dimethyl sulfoxide (DMSO), hexamethylphosphoramide (HMPA), *N*-methyl-pyrrolidone (NMP), and acetonitrile that can hydrogen bond or complex with acetylene and acetylenic compounds; used to dissolve and activate acetylene.

Carbonylation Reaction of acetylene with carbon monoxide and alcohols to form acrylate esters; Reppe process.

Commodity chemicals Large-volume, multitonnage chemicals, some of which are derived from acetylene.

Ethynylation Reaction of acetylene with aldehydes and ketones to form acetylenic alcohols and diols.

Grignard Organomagnesium halide used in acetylenic and other syntheses.

Oil-well acidizing Use of acetylenic alcohols (alkynols) as corrosion inhibitors to protect steel pipe during acidizing operations undertaken to free oil from limestone formations.

Pesticides Agricultural chemicals including herbicides, insecticides, miticides, fungicides, and bacterial control agents.

Reppe products and technology Pioneered by Dr. Walter Reppe of the I. G. Farben; various products derived from the reaction of acetylene with

formaldehyde to yield butyne-1,4-diol and propargyl alcohol.

Surfynols Acetylenic hydroxyl compounds used as high-speed, low-foam wetting–dispersing agents and as agricultural adjuvants and formulation aids.

Trofimov reaction Formation of substituted pyrroles and *N*-vinylpyrroles by reaction of acetylene with ketoximes.

Vinylation Reaction of acetylene with alcohols or pyrollidone to form vinyl ethers and vinylpyrrolidone.

ACETYLENE CHEMISTRY is the chemistry of the carbon–carbon triple bond (—C≡C—). This functionality defines the unique chemistry of this reactive group, in addition to its diverse and important applications. The high electron density of the triple bond with its circular, symmetrical π field makes acetylene and its derivatives reactive and useful intermediates for synthesizing a wide variety of organic compounds. These organic products find wide use in the synthesis of flavors and fragrances, vitamins A, E, and K, β-carotene, pesticides, surfactants, corrosion inhibitors, and specially intermediates. This article describes the technology and applications of acetylene, acetylenic compounds, and the chemicals derived from them.

In the mid-1960s more than 1 billion pounds of acetylene were used annually for the production of large-volume (commodity) chemicals. Since then, acetylene has been gradually supplanted by less expensive olefin feedstocks. However, acetylene is still used in multimillion-pound levels to produce Reppe chemicals (butynediol, propargyl alcohol, butanediol, butyrolactone, *N*-methylpyrrolidone, polyvinylpyrrolidone, and vinyl ether copolymers) and specialty acetylenic chemicals and their derivatives. Large volumes of butanediol are used in the manufacture of engineering plastics such as polybutylene terephthalate. Other significant uses for acetylene in specialty areas include acetylene black, vitamins A and E, flavor–fragrance (F & F) compounds, corrosion inhibitors, acetylenic surfactants, and pesticides. Acetylenic chemicals, polymers, and derivatives of potential value in research and commerce are also discussed. Some special aspects of acetylene chemistry research in Russia are also summarized.

I. INTRODUCTION

Acetylene has always been an important raw material for making chemical products. In the early years of its history (circa 1890–1900) it was used extensively as an illuminant for trains and city streets. It was soon realized that acetylene burned in the presence of oxygen provided a very hot flame, useful in the joining of metals. The welding industry today still uses significant amounts of acetylene in spite of the availability of less expensive fuels such as propane.

Initially, acetylene was handled in industry as the undiluted liquefied gas below its critical temperature of 36°C at a pressure greater than 600 psig. This appeared to be a safe procedure until a series of industrial explosions in the early 1900s eliminated the practice. Today acetylene is shipped in cylinders under pressure that contain a mixture of diatomaceous earth or asbestos, acetone, and stabilizers. Another safe method of transport is via granular calcium carbide in sealed containers, free of water. The large-volume use of acetylene for the manufacture of commodity chemicals has led to the building of plants, either petrochemical or calcium carbide, "across the fence" from acetylene producers. Today the factors leading to acetylene hazards are well understood and have been well documented. Acetylene now poses a minimum risk in well-operated processes and plants.

II. ACETYLENE AND COMMODITY CHEMICALS

From 1965 to 1970 more than 1 billion pounds of acetylene were used annually to manufacture a variety of chemicals. Welding applications constituted ~10% of this total. After 1970, the use of acetylene for the manufacture of commodity chemicals began to decline markedly, and by 1979–1983 only 269 million pounds were employed. Less expensive petrochemical raw materials such as ethylene, propylene, butadiene, amylenes, and methane were replacing acetylene. Table I summarizes chemicals manufactured mainly from actylene before 1965, types of processes, and replacement raw materials. The principal use of the monomers listed in the table was in diverse polymer applications spanning plastics, latex emulsions, rubbers, and resins. Chlorinated solvents were important in vapor degreasing, but their use in more recent years has been gradually limited as a result of toxicity and air pollution.

The applications of the polymer products derived from the above monomers are not within the scope of this article.

III. PRODUCTION OF ACETYLENE AND COMMODITY CHEMICALS

The following tabulation shows the gradual decline in U.S. acetylene production from its high point of 1.23 billion pounds in 1965 to ~269 million pounds in 1979. In 1965 bulk acetylene was valued at 7–12¢/lb, while ethylene was

TABLE I Early Acetylene-Based Chemicals

Product	C_2H_2 process	Replacement raw material	Replacement process
Acrylates and acrylic acid	Reppe carbonylation ($CO + C_2H_2$)	Propylene (C_3H_6)	Two-stage oxidation
Acrylonitrile	$C_2H_2 + HCN$	C_3H_6	Ammoxidation (C_3H_6–O_2–NH_3)
Chloroprene	C_2H_2-Vinylacetylene-HCl	Butadiene	Chlorination and dehydrochlorination
Chlorinated hydrocarbons	$C_2H_2 + Cl_2$	C_1–C_3 feedstocks; C_2H_4	Chlorination–dehydrochlorination
Vinyl acetate	C_2H_2 + acetic acid	Ethylene (C_2H_4)	Oxyacetylation
Vinyl chloride	$C_2H_2 + HCl$	Ethylene	Oxychlorination
	$C_2H_2 + HCl$	$C_2H_2 + C_2H_4$	Balanced ethylene–acetylene

3–4¢/lb. By 1983–1984 the cost ratio was approximately the same, with acetylene valued at about 55–75¢/lb and ethylene at 23–29¢/lb.

Year	C_2H_2 used (10^6 lb)
1965	1230
1967	1065
1969	1195
1971	852
1973	571
1976	490
1979	269
1984	286

From 1967 to 1974, 23 plants making such acetylene-based products as acrylonitrile, chlorinated hydrocarbons, chloroprene (neoprene), vinyl acetate, and vinyl chloride were shut down. Sixteen of these plants manufactured vinyl acetate and vinyl chloride. Table II presents acetylene usage for various products in 1970, 1979, and 1984.

In 1984 total U.S. capacity for the production of acetylene was estimated to be 384 million pounds. This production is a mix from calcium carbide, by-product acetylene from cracking, and partial oxidation processes.

TABLE II U.S Acetylene Usage: Large-Volume Chemicals

Product	1970	1979	1984
Acrylic acid and acrylates	70	16–45[a]	0
Acrylonitrile	42	0	0
Chloroprene (neoprene)	242	0	0
Chlorinated solvents	91	0	0
Vinyl chloride	268	100–110[a]	146
Vinyl acetate	158	37–52[a]	10
Reppe chemicals[b]	41	73–80[a]	114
Other acetylenics and derivatives[c]	10	14	26

[a] Mainly butane-1,4-diol plus other Reppe products.
[b] Acetylene black, vinyl fluoride, specialty acetylenics.
[c] Estimated value.

The nine U.S. acetylene producers, with their capacity in millions of pounds, were AIRCO-BOC, Calvert City, KY, and Louisville, KY (75); Dow, Freeport, TX (16); Hoffmann–La Roche, Nutley, NJ (5); Monochem, Geismar, LA (180); Rohm and Haas, Deer Park, TX (55); Union Carbide, Ponce, P. R. (12); Union Carbide; Seadrift, TX (12); Taft, LA (10); Texas City, TX (16).

The 1984 demand for acetylene was 286 million pounds, and it was estimated to be 292 million pounds in 1988. Growth from 1974 to 1983 was negative at −6.9% per year, while through 1988 it was slightly positive at 0.5% per year. Hoffmann–La Roche generates acetylene from calcium carbide for use in the manufacture of vitamins A and E and β-carotene.

Of all the commodity chemicals listed in Table II, vinyl chloride showed the least decline from 1970 to 1984. In 1984 acetylene converted to vinyl chloride represented 51% of total acetylene consumption. However, this production was from only one site, Monochem at Geismar, LA, and may be vulnerable in the future if Monochem decides to convert completely to ethylene as raw material. The most promising growth areas for acetylene in the near term are Reppe chemicals, particularly butane-1,4-diol, used extensively in engineering plastics and polyurethanes. Acetylene black and vinyl fluoride are also specialty growth areas, as are acetylenic surfactants and corrosion inhibitors. These acetylene-drived products are discussed in greater detail in Section V.

A. Acetylene Production on a World Basis

Table III shows that, although U.S. acetylene production is modest, acetylene usage worldwide is still significant, amounting to ∼1.9 billion pounds.

In the longer term, it is believed that worldwide acetylene capacity and usage will gradually increase as oil prices escalate. Acetylene usage for such products as vinyl acetate, vinyl chloride, Reppe chemicals, and specialty

TABLE III World Acetylene Usage[a]

Location	Chemical	Industrial	Total
United States	282	114	396
Western Europe	814	200[b]	1014[b]
Japan	106[b]	100[b]	206
Others	200[b]	150[b]	350[b]
Total	1402[b]	564[b]	1966[b]

[a] In millions of pounds.
[b] Estimated value.

acetylenics is expected to increase. Worldwide acetylene capacity is spread over a wide geographic area, as shown in Table IV. The calcium carbide (CaC_2) process, based on coal and limestone, is still extensively practiced or is present as a backup capacity. In Russia there is probably a large calcium carbide capacity that has not been reported.

The large-scale use of acetylene for the manufacture of commodity chemicals will be dependent on the cost difference between coal and oil and natural gas. It is certain that sometime in the future oil and natural gas reserves will become limited and more expensive than coal. The target date is the early 21st century. Coal-based technologies such as the calcium carbide and AVCO (coal–hydrogen plasma arc) processes are prime candidates for large-scale acetylene production. The AVCO process (Section VIII) has been studied successfully at the pilot-plant level.

IV. ACETYLENE-BASED PROCESSES FOR LARGE-VOLUME CHEMICALS

The processes summarized in the equations below were important in 1940–1965 for producing commodity chemicals. Below each acetylene-based process is shown its replacement process.

A. Acrylates and Acrylic Acid

1. Reppe Carbonylation

$$4C_2H_2 + Ni(CO)_4 + 4C_2H_5OH + 2HCl \longrightarrow CH_2{=}CHCO_2C_2H_5 + H_2 + NiCl_2$$

$$CH_2{=}CHCO_2C_2H_5 \xrightarrow{H_2O} CH_2{=}CHCO_2H + C_2H_5OH$$

$t = 30$–$40^\circ C$; atmospheric pressure.

2. Replacement Process: Two-Stage Propylene Oxidation

$$CH_3{-}CH{=}CH_2 + \tfrac{1}{2}O_2 \xrightarrow[300-450^\circ C]{A}$$

$$CH_2{=}CHCHO \xrightarrow[275-365^\circ C]{B} CH_2{=}CHCO_2H \xrightarrow{ROH}$$

Acrylate esters

A and B = fixed or fluidized-bed reactors.

B. Acrylonitrile

1. Acetylene–Hydrogen Cyanide

$$C_2H_2 + HCN \xrightarrow[40-600^\circ C]{} H_2C{=}CHCN$$

Fixed-bed process; catalyst, $Ca(CN)_2$.

2. Replacement Process: Ammoxidation of Propylene

$$H_3C{-}CH{=}CH_2 + \tfrac{3}{2}O_2 + NH_3 \rightarrow H_2C{=}CHCN + 3H_2O$$

TABLE IV Worldwide Acetylene Capacity

Country (Company, Location)	Capacity (millions of pounds)	Acetylene process
West Germany (BASF, Ludwigshaven)	176	Partial oxidation of natural gas
West Germany (Chem. Werke Huels, Marl)	264	Arc process–refinery gas
West Germany (BASF, Ludwigshaven)	13	By-product C_2H_2 from ethylene
Italy (Anic, Ravenna)	132	Naphtha cracking
Italy (Montedison, Porto Marghera)	154	Partial oxidation of natural gas
Japan (Denki Kagaku Kagyo, Ohmi)	200	Calcium carbide process
Japan (Igegana Electric Co., Ogaki)	616	Calcium carbide process
South Africa (African Explosives)	110	Calcium carbide process
Russia (Lissit Chansk)	77	Partial oxidation of natural gas
United States (Kentucky, Texas)	384	Calcium carbide, cracking by-product, partial oxidation

Fixed-bed or fluidized-bed reactors; $t = 240$–$460^\circ C$; typical catalysts, Bi–P–Mo (Sohio), Mo–Vo–Bi (Snam), Mo–Te–Ce (Montedison–UOP), Se–CuO (BP-Dist.)

C. Chloroprene (Neoprene, 2-Chloro-1,3-Butadiene)

1. Acetylene–Vinylacetylene–Hydrogen Chloride (Nieuwland–Carothers Process)

$$2C_2H_2 \xrightarrow[NH_4Cl]{CuCl} H_2C{=}CH{-}C{\equiv}CH \xrightarrow[HCl]{} H_2C{=}CH{-}CCl{=}CH_2 + HCl$$

Liquid-phase process; $t = 50$–$75^\circ C$; atmospheric pressure.

2. Replacement Process: High-Temperature Chlorination of Butadiene

$$Cl_2 \xrightarrow{\Delta} 2Cl\cdot$$

$$H_2C{=}CH{-}CH{=}CH_2 + 2Cl\cdot \xrightarrow{600^\circ C} H_2C{=}CH{-}CCl{=}CH_2 + HCl$$

D. Chlorinated Hydrocarbons (Solvents)

1. Chlorination of Acetylene Followed by Dehydrochlorination

$$C_2H_2 + Cl_2 \rightarrow CHCl_2CHCl_2 \xrightarrow{-HCl} CHCl{=}CCl_2 \xrightarrow[Cl_2]{} CHCl_2CCl_3 \xrightarrow{\Delta} Cl_2C{=}CCl_2$$

2. Replacement Process

This includes the chlorination of hydrocarbon feedstocks (methane, ethane, propane, ethylene, propylene), which requires multiproduct technology.

E. Vinyl Acetate

1. Acetylene–Acetic Acid

$$C_2H_2 + CH_3CO_2H \xrightarrow[180-210^\circ C]{} H_2C{=}CHOCOCH_3 + H_2O$$

Fixed-bed reactor; catalyst, zinc acetate on carbon.

2. Replacement Process: Oxyacetylation of Ethylene

$$H_2C{=}CH_2 + CH_3CO_2H + \tfrac{1}{2}O_2 \xrightarrow[175-200^\circ C]{} H_2C{=}CHOCOCH_3 + H_2O$$

Fixed-bed or fluidized-bed reactors; catalyst, palladium or mixture of noble-metal salts deposited on supports.

F. Vinyl Chloride

1. Acetylene–Hydrogen Chloride (Hydrochlorination)

$$C_2H_2 + HCl \xrightarrow[90-140^\circ C]{} H_2C{=}CHCl$$

Fixed-bed reactor; catalyst, 10% mercuric chloride on carbon.

2. Balanced Ethylene–Acetylene (Chlorination–Dehydrochlorination)

$$C_2H_4 + Cl_2 \rightarrow ClCH_2CH_2Cl \rightarrow H_2C{=}CHCl + HCl$$

$$C_2H_2 + HCl \rightarrow H_2C{=}CHCl$$

$$\text{Overall: } C_2H_4 + C_2H_2 + Cl_2 \rightarrow 2H_2C{=}CHCl$$

3. Replacement Process: Oxychlorination of Ethylene

$$C_2H_4 + 2HCl + \tfrac{1}{2}O_2 \rightarrow ClCH_2CH_2Cl + H_2O$$

$$C_2H_4 + Cl_2 \rightarrow ClCH_2CH_2Cl$$

$$ClCH_2CH_2Cl \xrightarrow{\Delta} H_2C{=}CHCl + HCl$$

$$2HCl + \tfrac{1}{2}O_2 \rightarrow Cl_2 + H_2O$$

$$\text{Overall: } 2C_2H_4 + Cl_2 + \tfrac{1}{2}O_2 \rightarrow 2H_2C{=}CHCl + H_2O$$

Fixed-bed or fluidized-bed reactors; oxychlorination, $t = 450$–$500^\circ C$ (atmospheric pressure); catalyst, $CuCl_2$–KCl on Kieselguhr; chlorination of ethylene, $t = 50$–$140^\circ C$ ($P = 4$–10 atm); pyrolysis of $ClCH_2CH_2Cl$, $t = 470$–$540^\circ C$ ($P = 24$–25 atm).

These acetylene-based processes are well-tested, high-yield, and high-selectivity operations, averaging well over 90% of theory. The extensive use of these processes is likely once oil-based feedstocks become more expensive than coal, which will be the preferred raw material for acetylene production.

V. IMPORTANT CHEMICAL USES FOR ACETYLENE

As shown in Tables II and III, there has been steady growth in the production of acetylene-based specialty chemicals. These products comprise acetylene black, vinyl fluoride, vinyl ethers, Reppe chemicals, and miscellaneous acetylenic alcohols and diols. Reppe products consist mainly of butyne-1,4-diol, propargyl alcohol (1-propyn-3-ol), butane-1,4-diol, butyrolactone, tetrahydrofuran, pyrrolidone, vinylpyrrolidone, *N*-methylpyrrolidone, polyvinylpyrrolidone, polyvinyl ethers, and copolymers. Butane-1,4-diol has emerged as the most important user of acetylene for acetylenic chemicals and derived products. Its usage in specialty polymers and chemicals increased steadily during the 1977–1979 period from 82 to 92% of total acetylene consumption (74 million pounds in 1979). The principal uses for butanediol are in polybutylene terphthalate and other engineering plastics, besides polyurethanes. By 1984 the total acetylene usage for acetylenic chemicals had been estimated to be 115 million pounds.

A. Acetylene Black

This unique, conductive form of carbon is made by the exothermic decomposition of acetylene:

$$C_2H_2 \xrightarrow{800^\circ C} 2C + H_2.$$

The reaction is highly exothermic and the evolved heat (∼55,000 cal/g) is used to maintain the pyrolysis temperature. Preferred processes involve (1) continuous explosion of acetylene at 1–2 atm using an electric spark and (2) decomposition via the electric arc.

The principal uses of acetylene black are in the manufacture of dry cell batteries and the fabrication of conducting rubber and plastic sheeting and molded plastics. Although acetylene black is too expensive to compete with petrochemical carbons in large-volume polymer compounding, it has carved out for itself a growing specialty market in conductive applications. Acetylene consumption (in estimated millions of pounds) for acetylene black grew steadily from 1967 to 1984 in the United States: 1967 (10); 1979 (19); 1980 (20); 1984 (23).

Producers of acetylene black are Shawinigan (Canada), Union Carbide (Ashtabula, OH, and Ponce, P. R.), and Gulf (Cedar Bayou, TX). The Gulf production facility went on stream in 1979 and is rated at 15–20 million pounds/year. The markets for batteries and conducting polymer composites grew significantly from the mid-1980s, and acetylene black participated in this growth. The total production of acetylene black from all sources (millions of pounds) from 1981 to 1987 is estimated as follows: 1981 (50); 1984 (55); 1987 (60). Union Carbide produced ∼65% of this total, which was used internally in its battery division.

B. Vinyl Fluoride

The principal use for this specialty monomer is the production of polyvinyl fluoride (PVF), a polymer considerably more stable than polyvinyl chloride (PVC). The use of PVF in exterior coatings for aluminum siding, steel building panels, hardboard siding, and asbestos-impregnated felt roofing continues to grow. PVF was introduced by Du Pont in 1963 under the registered trademark Tedlar. In 1966 acetylene consumption for PVF production was estimated to be ∼1 million pounds and by 1984 it was 6–9 million pounds, showing steady growth.

Vinyl fluoride is produced by the addition of hydrogen fluoride to acetylene in the presence of a mercuric salt catalyst deposited on carbon. The intermediate vinyl fluoride is not isolated due to the difficulty of separating it from acetylene. The resulting difluoroethane (DFE) is in turn cracked to vinyl fluoride and the evolved hydrogen fluoride is recycled:

$$C_2H_2 + 2HF \rightarrow [\underset{\text{Vinyl fluoride}}{H_2C{=}CHF}] \rightarrow \underset{\text{DFE}}{CH_3CHF_2}$$

$$DFE \xrightarrow{\Delta} H_2C{=}CHF + HF.$$

The principal U.S. producers of vinyl fluoride and PVF are Du Pont (Louisville, KY) and Diamond Shamrock (Houston, TX). The intermediate DFE is also important for the manufacture of vinylidene fluoride, as shown in the next section.

C. Vinylidene Fluoride (Vinylidene Difluoride)

Vinylidene fluoride (VDF), a monomer, can be produced from DFE via chlorination and cracking or from 1,1,1-trifluoroethane:

$$C_2H_2 + 2HF \rightarrow CH_3CHF_2 \xrightarrow{Cl_2}$$

$$\underset{\text{CDFE}}{H_3C{-}C(Cl)F_2} \xrightarrow{\Delta} \underset{\text{VDF}}{H_2C{=}CF_2} \qquad (1)$$

$$CH_3CF_3 \xrightarrow{\Delta} H_2C{=}CF_2 + HF \qquad (2)$$

At present the preferred starting material for VDF production is 1-chloro-1,1-difluoroethane (CDFE), also known as refrigerant 142b. Numerous other intermediates have been described for the preparation of VDF. The principal use for this monomer is the production of polyvinylene difluoride (PVDF, PVF_2). The polymer is semicrystalline and has the following desirable properties; high mechanical and impact strength, resistance to most chemicals, very high dielectric constant, and excellent resistance

to ultraviolet and nuclear radiation and general weathering. It can be cross-linked via electron-beam radiation to give further improved properties.

It is used extensively in the interior coating of pipes and as a base for long-lasting decorative finishes on aluminum, galvanized steel, and other architectural metals. Other applications include wire insulation and connectors for computer boards, telecommunications uses, and resistant coatings for harsh environments.

The U.S. suppliers of PVDF are Pennwalt, Solvay, and Dynamit Nobel. Du Pont produces PVDF for its own internal use, primarily to produce Viton plastic. The total U.S. production of this polymer is significantly greater than that of PVF.

VI. REPPE PRODUCTS

Reppe products, a line of large-volume, acetylene-based specialty chemicals, bear the name of their discoverer, Walter Reppe of the I. G. Farben, who is one of the legendary figures of acetylene chemistry and its application to the production of valuable chemicals for Germany during World War II. Reppe's research spanned the 1930s and reached commercial importance during the war years in the 1940s. Ethynylation technology to produce the key starting materials 2-butyne-1,4-diol and propargyl alcohol (1-propyn-3-ol) is summarized in Section IX.B.

$$C_2H_2 + HCHO \xrightarrow{CuC_2H} \underset{\text{2-Butyne-1,4-diol}}{HOCH_2{-}C{\equiv}CCH_2OH} + \underset{\text{Propargyl alcohol}}{HOCH_2C{\equiv}CH}$$

Products derived from butynediol are 2-butene-1,4-diol, 1,4-butanediol, γ-butyrolactone, *N*-methyl-2-pyrrolidone, 2-pyrrolidone, *N*-vinyl-2-pyrrolidone, polyvinylpyrrolidone (PVP), and polypyrrolidone. The reaction flow diagram in Fig. 1 shows the interrelationship of these products. Other Reppe chemicals such as vinyl ethers, polyvinyl ethers, vinyl ether–maleic anhydride

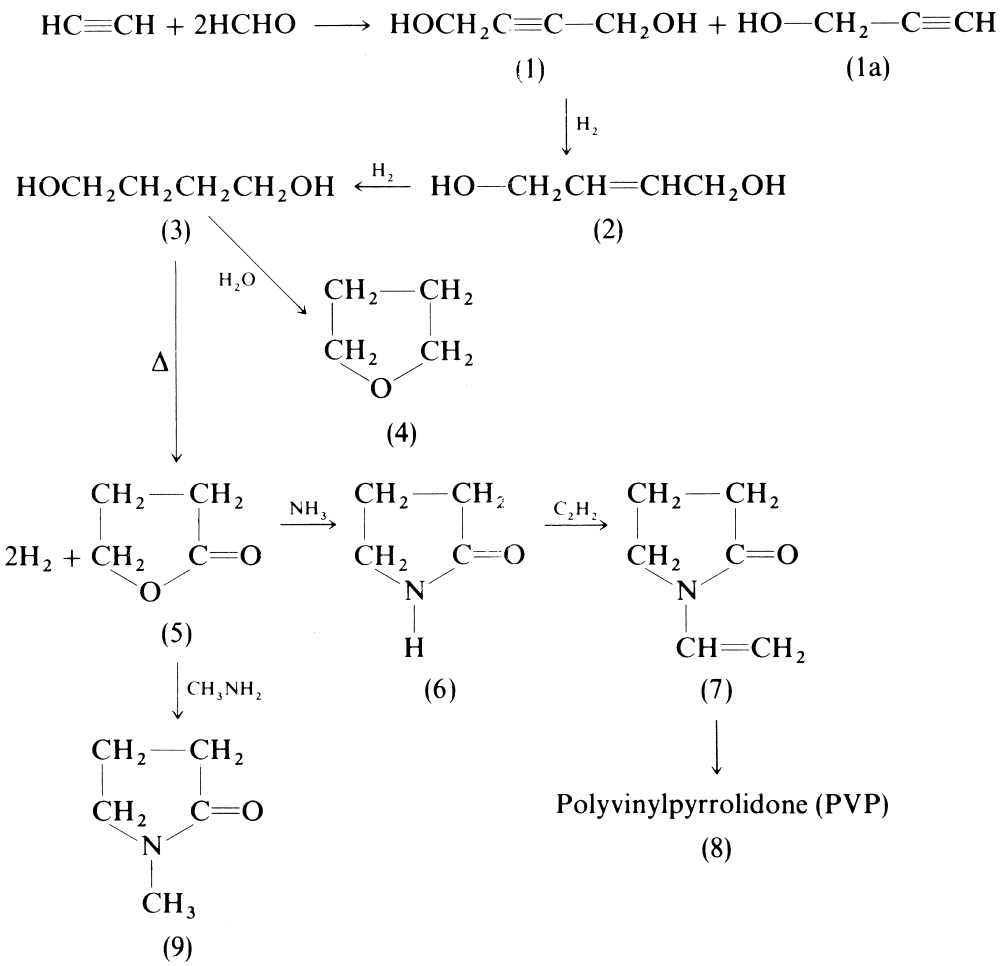


FIGURE 1 Reaction flow diagram: Reppe chemicals. (1) 2-Butyne-1,4-diol; (1a) propargyl alcohol; (2) 2-butene-1,4-diol; (3) butane-1,4-diol; (4) tetrahydrofuran; (5) γ-butyrolactone; (6) 2-pyrrolidone; (7) *N*-vinylpyrrolidone; (8) polyvinylpyrrolidone; (9) *N*-methylpyrrolidone.

TABLE V Applications of Reppe Chemicals[a]

Product	Applications[b]
Propargyl alcohol	Metal treatment; corrosion inhibition; oil-well acidizing; electroplating; intermediate for vitamin A and pesticides
Butyne-1,4-diol	Important Reppe starting material; metal treatment; acid pickling; electroplating additive; stabilization of chlorinated solvents; intermediate for pesticides
Butene-1,4-diol	Intermediate for pesticides, pharmaceuticals, fungicides, and bacteriacides; polyurethane intermediate
Butane-1,4-diol	Important Reppe intermediate for THF, butyrolactone, and PVP polymers; manufacture of polybutylene terphthalate; polyurethanes, spandex fibers; specialty plasticizers
γ-Butyrolactone	Intermediate for pyrrolidone and PVP; acetylene solvent (Sacchse); specialty solvent for polymers, lacquers, paint removers, and petroleum processing; intermediate for herbicides, azo dyes, and methionine
Tetrahydrofuran	Continuous top coating of automotive vinyl upholstery, general solvent for resins; coating of cellophane with vinylidene polymers; polymers; polyurethane polymers; spandex fibers
Pyrrolidone	Intermediate for *N*-vinyl and PVP polymers; powerful solvent for resins; acetylene solvent; polar reaction solvent; formulating agent; floor waxes; specialty inks; reaction intermediate
N-Methylpyrrolidone	Commercial acetylene solvent; uses same as pyrrolidone; selective extraction solvent for butadiene purification (cracked naphtha); spinning synthetic fibers; surface coatings; pigment dispersant; formulating agent
N-Vinylpyrrolidone	Intermediate for PVP polymers; functional monomer; lube oil manufacture; copolymer applications; paints, paper-coating adhesives, cosmetics; increased dye receptivity; pigment dispersant

[a] Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," pp. 101–102, courtesy of Dekker, New York.
[b] THF, tetrahydrofuran; PVP, polyvinylpyrrolidone.

copolymers, acetylenic alcohols and diols, and polybutylene terphthalate are discussed in subsequent sections.

A. Applications of Reppe Products

The uses of both butynediol and propargyl alcohol and their derivatives are quite diverse (Table V). They vary from metal treatment and corrosion inhibition to intermediates, specialty solvents for acetylene and polymers, top coating and spinning of fibers, preparation of specialty polymers and engineering plastics, vitamins A and E, and pesticides. The most important products in terms of production and sales are butanediol and tetrahydrofuran (THF), which are key building blocks for engineering plastics, polyurethanes, and spandex fibers. Table VI

TABLE VI Applications of Polyvinylpyrrolidone Polymers[a]

Product	Composition[b]	Applications
Polyvinylpyrrolidone	—	Approved by the Food and Drug Administration for food and drug use; blood plasma expander; nontoxic, biodegradable; varied uses
Polyclar AT	Standard PVP grade (white powder)	Beverage clarification aid (fruit juices, wines, beer); chill proofing of beer; complexing agent
Plasdone	Pharmaceutical-grade PVP	Tablet manufacture (granulating agent); tablet coating; liquid dosages; topical preparations; stabilizer, dispersant, drainage aid (syringes); skin creams, hair sprays, shampoos, general cosmetic use
Kolima adhesive polymers (Kolima 35, 55, 75)	Modified, clear high-solids PVP solutions	Superior glue-line strength; sealing of smooth hard surfaces; grease and chemical resistance; superior adhesion at temperature extremes; superior surface wetting; residual tack and viscosity control applications
Ganex V polymers	Proprietary PVP polymers	Coatings, detergents, pigment dispersants, plastic additives, textiles, petroleum applications, protective colloids for vinyl lattices, improved freeze-thaw stability and scrub resistance
Polectron (P) emulsion copolymers	PVP copolymer lattices: P-130 (ethyl acrylate); P-230 (2-ethylhexylacrylate); P-430, 450 (styrene); P825L, 845L (vinyl acetate)	Adhesives: remoistenables, pressure sensitise, heat-sealing applications Coatings: leather sizes and finishes, metal primers Paper: board sizing, pigment binding, precoating, heat sealing Textiles: fabric laminates, oil repellent finishes, permanent pressing, polyester sizing

[a] Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and other Natural Resources," pp. 134–135, courtesy of Dekker, New York.
[b]

TABLE VII Applications of Vinyl Ethers and Polymers[a]

Product	Composition	Applications
Methyl vinyl ether	$CH_3O{-}CH{=}CH_2$	Specialty monomer; intermediate; Gantrez resins
Ethyl, *n*-butyl, isobutyl, decyl vinyl ethers	CH_2H_2 + alcohol $\rightarrow$ $RO{-}CH{=}CH_2$	Specialty monomers; copolymers (adhesives); lacquers, paints, plasticizers, pour-point depressants; reaction intermediates
Gantrez M	Polyvinyl methyl ether grades M-55, M-154 (50% H_2O) M-574 (70% solids in toluene)	Adhesive formulations with high wet tack; adhesion to plastic and metal surfaces; uses: adhesives, coatings, printing inks, textiles
Gantrez AN	Copolymer of methyl vinyl ether and maleic anhydride grades AN-139, AN-149, AN-169	Polyelectrolyte in aqueous media; uses: hair sprays, shampoors, detergents, leather, paper coating, pigment grinding, nonwoven fabrics, topical remedies, tablet coating, photoreproduction
Gantrez VC	Copolymers of alkyl vinyl ethers (e.g., isobutyl vinyl ether) and vinyl chloride	Traffic, marine, and maintenance surface coatings; foundation sealants; corrosion-resistant paints; flame retardancy for vinyl substrates; primer coat for metals, paper coating

[a] Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 142, courtesy of Dekker, New York.

summarizes typical uses of polymers and copolymers derived from *N*-vinylpyrrolidone.

Propargyl alcohol is used extensively in the United States as a corrosion inhibitor in acid media for such applications as oil-well acidizing and acid pickling of steel and electroplating and as an intermediate for vitamin A and pesticides. These combined uses are estimated to comprise about 3–5 million pounds annually.

Other Reppe derivatives such as γ-butyrolactone, pyrrolidone, *N*-methylpyrrolidone, and *N*-vinylpyrroli done are used mainly as intermediates in the manufacture of PVP. However, butyrolactone and *N*-methylpyrrolidone are used extensively as solvents for polymers and acetylene. Butynediol has minor uses in metal treatment and as a pesticide intermediate (Barban). However, it is very important as a starting point in the manufacture of THF and butanediol, the two largest volume Reppe products. Both of these products have large-volume uses in the manufacture of engineering plastics and polyurethane polymers (see Table V and Sections VI.C and VI.D). PVP has important uses in the clarification of beer, wine, and fruit juices, in pharmaceutical tablet manufacture, and in cosmetic manufacture. Its high degree of solubility and compatibility makes it useful in numerous adhesive applications.

Vinyl ethers and derived polymers are summarized in Table VII. These products are particularly useful in formulating adhesives and coating additives. The copolymer of methyl vinyl ether and maleic anhydride is sold by General Aniline and Film Corporation (GAF) under the name Gantrez AN series. These copolymers are important in hair sprays and shampoos and compete with the PVP polymers as versatile specialty polymers. Polyvinyl methyl ether and copolymers have various uses as adhesives, particularly in high-tack formulations.

Vinyl ether monomers are prepared by the base-catalyzed reaction of alcohols with acetylene. With reactive alcohols such as methanol, ethanol, and isopropanol, the reaction can be carried out at atmospheric pressure:

$$C_2H_2 + ROH \rightarrow RO{-}CH{=}CH_2.$$

The original Reppe process involved heating the alcohol–acetylene charge at 150–180°C using 1–2% KOH or sodium as catalyst under pressure, with a partial nitrogen atmosphere. GAF has produced a variety of alkyl vinyl ethers, but the most important is methyl vinyl ether, probably due to the commercial success of its copolymer series with maleic anhydride. The vinyl ethers polymerize readily via free-radical catalysts to yield both homo- and copolymers. They are reactive molecules and can also be used as intermediates to make a large variety of products.

B. Butane-1,4-Diol and Tetrahydrofuran: Large-Volume Polymer Uses

Butanediol has evolved into the most important chemical in the Reppe line on the basis of its uses in the manufacture of polybutylene terephthalate (PBT) and THF. It is also important as a chain extender for polyurethanes. If total acetylene usage for acetylenic chemicals in 1984 is estimated to be 115 million pounds and butanediol production is estimated conservatively to be 75% of this total, then at least 270 million pounds of butanediol were then produced.

C. Polybutylene Terephthalate

Although this engineering plastic competes with polyethylene terephthalate (PET), PBT is considered a superior polymer because of its outstanding properties. It is produced by ester interchange between dimethyl terephthalate and butanediol, while PET is derived from ethylene glycol as the diol component. PBT is most widely used as a 30% glass-filled blend.

The desirable properties of PBT include excellent dimensional stability, superior strength and stiffness, exceptional moldability and low mold shrinkage, low creep at elevated temperatures, excellent electrical (dielectric) properties, and good chemical resistance.

It has end uses in the automotive industry: dielectric and heat-resistant uses to 300°F; ignition system components—distributor caps, coil bobbins, rotors; light sockets; emission control parts; body panels and louvers; transmission components. It is also used in electronics for connectors, switches, relays, cases, and covers. In the manufacture of appliances, PBT is a component of handles, small housings, impellors, and general engineering parts; it also has high-voltage applications (resistance to arc tracking).

The annual growth rate of PBT from 1974 to 1977 was ~40%, and in 1984 it was estimated to be 10–15%. If acetylene usage between butanediol and THF production is about equal, then PBT derived from acetylene could be of the order of 200 million pounds/year.

D. Tetrahydrofuran: Commercial Uses

Du Pont built a THF production facility in Houston in 1969 based on Reppe technology. The capacity of the plant was increased from 50 million to 90 million pounds annually by 1973. This plant made Du Pont the largest producer of THF in the United States. By virtue of the uses summarized below, THF has been growing at an annual rate of 7%; by 1979 U.S. production was ~112 million pounds.

1. Continuous top coating of automotive vinyl upholstery via THF–PVC solutions
2. Coating cellophane with vinylidene copolymers
3. Coating of PVC and acrylic polymer sheets
4. Protective coatings, film casting, printing inks, adhesives
5. Activating solvent for tetraethyllead manufacture, specialty solvent
6. Polymer intermediate for polyethers, polyurethane elastomers, spandex fibers

To date the Reppe process for both butanediol and THF remains dominant.

VII. SPECIALTY ACETYLENICS AND DERIVATIVES

Specialty acetylenics and their derivatives have a variety of uses which are described in the subsections that follow. Further details are provided in Sections IX–XII. The production volumes and dollar sales of such products are often proprietary and seldom exceed 2–15 million pounds.

A. Vitamins A, E, and K and β-Carotene

Acetylene is an important building block in the lengthy synthesis of vitamins A, E, and K and β-carotene, all of which are important polyterpene compounds. Synthetic vitamin E, also known as *dl*-α-tocopherol, has emerged as one of the most important vitamins in the world today. Before 1950 its role in nutrition was in dispute, and it was not considered an essential vitamin. The synthesis, development, and manufacture of vitamins A and E were pioneered by Hoffmann–La Roche in Basel, Switzerland and Nutley, NJ. Both vitamins are important in animal, human, and pet nutrition. The largest commercial markets are animal feeds and supplements. In 1983 the production of both synthetic and natural vitamin E was estimated to be 11 million pounds with a value of about $164 million. Vitamin A production is about one-tenth that of vitamin E, and that of vitamin K is even less. Vitamin K is used almost entirely as an antihemorrhagic agent, while β-carotene made from vitamin A (*trans*-retinol) is used as a safe food colorant. Since β-carotene is metabolized in mammalian organisms partly to vitamin A, it is far more effective than ordinary yellow food dyes. The technology of vitamins A and E and flavor–fragrance (F & F) compounds is summarized in Section XI.

B. Flavors and Fragrance Compounds

Although it is possible to produce a variety of F & F compounds by the reaction of acetylene with C_5 to C_8 ketones followed by hydrogenation, only a small number of these have attained commercial importance, as shown in the following subsections.

$$R_1R_2{-}C{=}O + C_2H_2 \xrightarrow[\text{solvent}]{\text{KOH}}$$

$$R_1R_2{-}\underset{\displaystyle OH}{\underset{|}{C}}{-}C{\equiv}CH \xrightarrow{H_2} R_1R_2{-}\underset{\displaystyle OH}{\underset{|}{C}}{-}CH{=}CH_2$$

1. Linalool (3,7-Dimethylocta-1,6-dien-3-ol)

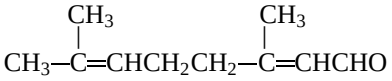
The fragrance compound linalool is prepared via the route shown above by the reaction of methylheptenone

(6-methylhept-5-en-2-one) with acetylene to yield dehydrolinalool (DHL, 3,7-dimethylocta-6-en-1-yn-3-ol), which is semihydrogenated to give the final product.

DHL is an important intermediate in the commercial synthesis of vitamin A, while linalool is a key intermediate in the production of vitamin E (*dl*-α-tocopherol). Linalool is one of the most important fragrance compounds in use today. It is used in many perfume compositions where a desirable floral-rose note is desired. In the past an important natural source of linalool was Brazilian rosewood. It has also had minor uses as a flavor ingredient. Besides the acetylenic route, large amounts of linalool are produced by a turpentine-based intermediate, myrcene. The Glidden–SCM Company produces a commercial grade of linalool that is used extensively as a vitamin E intermediate. The chemistry of DHL and linalool is summarized in Section XI.

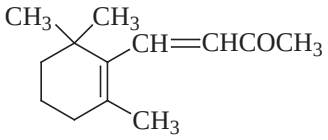
2. Citral (3,6-Dimethylocta-2,6-dienal)

The isomerization of the alkynol DHL with vanadium catalysts yields citral, an important lemon F & F compound used in many F & F compositions. Citral is also used as a vitamin A intermediate. Its condensation with acetone yields pseudoionone, which on cyclization (H_2SO_4) forms β-ionone, a key compound with the vitamin A ring system and a valuable perfumery compound (Section XI).



3. β-Ionone [4-(2,6,6-Trimethyl-1-cyclohexen-1-yl)-3-buten-2-one]

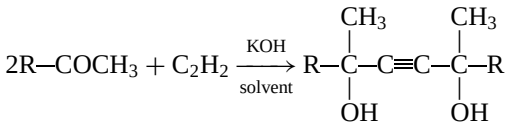
Besides the acetone–citral route just described, β-ionone is made by condensing the acetylenic alcohol DHL with diketene or methylisopropenyl ether. Both routes yield pseudoionone in comparable yields. The diketene route described in Section XI yields a tertiary acetylenic acetoacetate ester as the primary product, which is then rearranged with the loss of CO_2 to yield pseudoionone. The latter compound is readily converted to β-ionone, with dilute sulfuric acid as catalyst. This terpenic ketone, in addition to its use in the production of vitamin A, is an important ingredient of floral (iris) perfumes with woody notes. It also has minor uses as an ingredient in fruit flavors.



β-Ionone

C. Acetylenic Surfactants

The reaction of methyl ketones such as methyl ethyl and methyl isobutyl ketones with acetylene via base-catalyzed ethynylation yields tertiary acetylenic diols. These compounds containing the 1,4-dihydroxyacetylenic grouping at carbon chain lengths of eight or higher are superior



surfactants. They are produced by Air Products and Chemicals, Inc., and are sold under the trademark Surfynol™. Typical products are listed in the tabulation below:

Trademark	Generic name or description
Surfynol 82	3,6-Dimethyl-4-octyne-3,6-diol
Surfynol 104	2,4,7,9-Tetramethyl-5-decyne-4,7-diol
Surfynol 104-A	Surfynol 104 + 50% 2-ethylhexanol
Surfynol TG	Proprietary Surfynol surfactant
Surfynol 440	Ethoxylated (40%) Surfynol 104

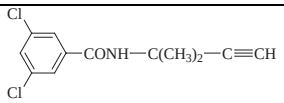
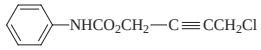
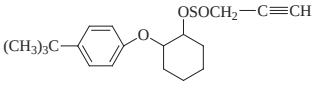
These surfactants, particularly Surfynols 104 and 440, are superior wetting–dispersing agents with very low foam characteristics. They are used extensively in water-based industrial primers and finishes, in rinse-aid applications, in pigment grinding, and in agricultural formulations. They give synergistic performance with other surfactants in terms of wetting and defoaming action, and in coating formulations they yield latex paints that can be painted directly on oily surfaces without the finished paint crawling or developing "fish eyes" on drying. These specialty surfactants have enjoyed steady growth since 1974, and a 1984–1985 production estimate was 5 million pounds annually.

D. Corrosion Inhibitors

Acetylenic alcohols have been used extensively since the mid-1950s in oil-well acidizing operations to free trapped oil in limestone formations. The acetylenic alcohol (alkynol) is formulated with 15 to 28% HCl and pumped downhole via steel tubing (N-80, J-55) to the limestone formation. Some acidizing operations attain depths of more than 20,000 ft with temperatures greater than 300–400°F. The alkynols, by virtue of complexing or reacting with the iron surface, minimize HCl corrosion and

Alkynol	Chemical structure
Propargyl alcohol	$HO{-}CH_2{-}C{\equiv}CH$
1-Hexyn-3-ol	$CH_3(CH_2)_2{-}CH(OH){-}C{\equiv}CH$
4-Ethyl-1-octyn-3-ol	$CH_3(CH_2)_3{-}CH(C_2H_5){-}CH(OH){-}C{\equiv}CH$
OW-1	Proprietary acetylenic inhibitor

TABLE VIII Acetylenic Pesticides

Name	Uses	Starting material	Pesticide structure
Kerb (pronamide)	Control of grasses in lettuce production; other crops: alfalfa, trefoil, endive, escarole	3-Methyl-1-butyn-3-ol	
Barban (carbyne)	Control of wild oats in production of spring wheat, durum wheat, barley, sugar beets, soybeans, lentils, flax	2-Butyne-1,4-diol	
Omite (propargite)	Acaricide with residual killing action; control of many varieties of mites; very effective against spider mites; used on fruit (apples, apricots, oranges, peaches, plums) and field crops (corn, potatoes)	Proporgyl alcohol	

extend the life of the tubing for years. These products have been recommended for the acid pickling of steel to remove mill scale. This use is smaller than oil-well acidizing.

The alkynols used in these applications are propargyl alcohol (1-propyn-3-ol), hexynol, ethyl octynol, and OW-1 (proprietary alkynol-based product) (see table below).

The most effective acetylenic inhibitors are secondary acetylenic alcohols of C_6–C_8 chain length, with the hydroxyl group α to the terminal ethynyl group. Propargyl alcohol, a primary alkynol, is somewhat less effective than the best inhibitors, but since it is a by-product of the Reppe process it is significantly less expensive than hexynol or ethyl octynol. The total 1984–1985 usage of acetylenic corrosion inhibitors was estimated to be about 3–4 million pounds with propargyl alcohol accounting for about half of this total.

E. Acetylenic Pesticides and Adjuvants

Acetylenic compounds have found selective uses as herbicides and miticides in the control of weeds and pests. It is likely that if more research and development effort were extended to triple-bonded compounds, superior pesticides would be developed. Three acetylenic compounds used commercially are Kerb (pronamide), Barban (carbyne), and Omite (propargite). They are summarized in Table VIII and Sections XII.A–XII.C in terms of the acetylenic starting material, pesticide structure, and uses.

It is believed that the use of Omite is considerably greater than that of either Kerb or Barban. Surfynol 104 (tetramethyldecynediol) and the Surfynol 400 series (ethoxylated S-104), while not pesticides, have been extensively evaluated as adjuvants for use with crop herbicides such as atrazine, alachlor, basagran, sencor, treflan, and kerb. In most cases they have been found superior to standard commercial surfactants of the nonionic and anionic types, if used with the above herbicides. They are also effective when used with bacterial control agents for the control of larva-worm infestations. They are used in significant amounts as agricultural formulation aids.

VIII. PROCESSES FOR ACETYLENE PRODUCTION

Common raw material sources for the production of acetylene are coal, natural gas, and liquid petroleum feedstocks such as naphthas and light oils. Calcium carbide technology using coal (coke) and limestone was the earliest means of producing acetylene. It is still used, but on a limited scale.

Petrochemical acetylene, based on natural gas and liquid feedstocks, began to emerge as a competitor of the carbide process in 1940–1950. Until the mid-1960s acetylene was still the favored raw material for such products as acrylates, vinyl chloride, vinyl acetate, chloroprene, acrylonitrile, and chlorinated hydrocarbons. However, as steam cracking of C_1–C_3 alkane feedstocks to olefins (ethylene, propylene) became important, the role of both petrochemical and carbide acetylene declined markedly by the mid-1970s. These olefin cracking processes also produced by-product acetylene, which could be recovered by selective solvent extraction and sold or utilized for chemical processes by nearby chemical companies. Companies such as Union Carbide, Tenneco, and Monochem have such by-product acetylene facilities in Texas and Louisiana.

Table IX summarizes processes that were or are practiced in the United States and elsewhere, with the exception of the newer AVCO coal-based plasma are process. Coal-based processes such as calcium carbide or AVCO will probably become important after the year 2000 as oil reserves become depleted.

Petrochemical acetylene processes are characterized by a common principle. A hydrocarbon feedstock is

TABLE IX Acetylene from Natural Gas, Petroleum, and Coal Sources[a]

Typical process	Principal feedstock	Technology	Typical companies
Electric arc	Methane, gas oils	Arc or plasma	Huels, Du Pont
Sachsse	Methane, natural gas mixtures	Partial combustion (one stage)	BASF, Dow, Monsanto
SBA	Methane, natural gas mixtures (first stage); naphthas, heavier feedstocks (second stage)	Partial combustion (one and two stage)	SBA, M. W. Kellogg
Wulff	Natural gas, naphthas, heavier feedstocks	Regenerative furnace pyrolysis (four cycles)	Union Carbide, Wulff
Montecatini	Natural gas, naphthas	Partial combustion under pressure	Montecatini, Diamond Alkali
By-product acetylene	Ethane, hydrocarbons, naphthas, oil	Steam cracking	Major oil and chemical companies (EXXON, Shell, Dow, Union Carbide)
Calcium carbide	Limestone ($CaCO_3$) and coke	CaC_2 from C + $CaCO_3$; C_2H_2 from CaC_2+ H_2O	AIRCO-BOC, Union Carbide
AVCO	Coal and hydrogen	Hydrogen plasma	AVCO (pilot and demonstration plants)

[a] Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 11, courtesy of Dekker, New York.

subjected to an intense energy source and thereby heated to 1200–1500 K. By the use of very short residence times (0.01–0.001 sec) and quick quenching of the cracked gas to 550 K, acetylene and the starting feedstock can be recovered. The recovery and purification of petrochemical acetylene is a lengthy operation compared with the simple calcium carbide process, which readily yields high-purity acetylene.

The petrochemical acetylene processes most likely to be practiced today are the partial oxidation types. Prominent among these are Sachsse, SBA, and Montecatini processes. The SBA and Montecatini processes utilize either natural gas or naphtha feedstocks, while the Sachsse process is designed primarily for natural gas or methane, but can be modified for naphtha. However, Sachsse technology has been widely practiced both in the United States and in Europe, showing it to be reliable and trouble free.

In contrast, the electric are process is more sensitive to process variables, which can lead to the formation of large amounts of by-product carbon. The Wulff process was once widely used in Europe, South America, Japan, and the United States. By controlling the feedstock and operating conditions, the four-cycle regenerative process could be made to deliver mainly acetylene or ethylene, making it more versatile than other petrochemical processes. The fact that the Wulff process is now seldom used may be due to the high efficiency of stream cracking of alkanes to form ethylene and propylene at lower energy usage, thereby making the Wulff mixed product stream less attractive.

A process diagram of the BASF Sachsse (burner) reactor is shown in Fig. 2. Normally, with natural gas the burner is a one-stage reactor. However, it can be modified to use naphtha by extension of the combustion chamber (second stage) and an oxygen–steam off-gas mixture to provide a moist flame zone in the lower end of the burner to crack naphtha to acetylene. About two-thirds of the hydrocarbon feed is burned in the reactor to provide the thermal energy needed to crack the remaining feed to acetylene.

The AVCO coal–hdrogen plasma process has not yet been scaled up to commercial production. However in a joint AVCO–DOE project (1980) at Wilmington, MA, a coal-fed hydrogen–plasma reactor capable of producing 2 million pounds/year of acetylene was successfully demonstrated, The pilot unit gave an acetylene yield of 35% based on coal with low electrical usage. AVCO claims

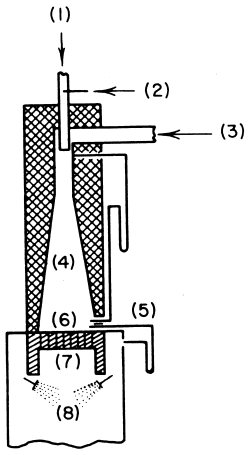


FIGURE 2 BASF (Sachsse) burner. (1) Oxygen; (2) inert gas (safety purge); (3) methane or naphtha feed; (4) neck and mixing chamber; (5) auxiliary oxygen; (6) burner block; (7) reaction chamber; (8) water quench. [Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 21, courtesy of Dekker, New York.]

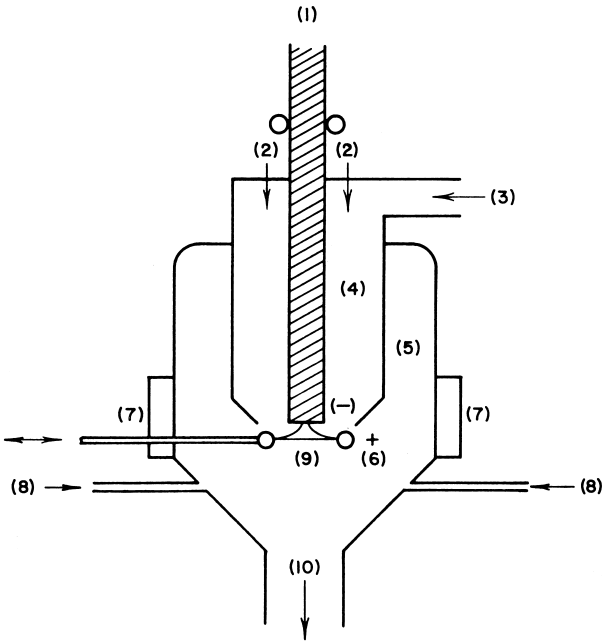


FIGURE 3 Coal–hydrogen (AVCO) plasma arc reactor. (1) Cathode and drive rolls; (2) hydrogen feed; (3) coal feed; (4) plenum; (5) reactor body; (6) anode; (7) field coils; (8) hydrogen quench ports; (9) rotating arc; (10) char and gas to cyclone. [Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 30, courtesy of Dekker, New York.]

that this process can be scaled rto commercial size to give acetylene commpetitive with or less expensive than ethylene used for commodity chemicals such as vinyl chloride and vinyl acetate.

Figure 3 is a schematic diagram of the AVCO plasma reactor. By means of an external magnetic field, the arc is made to rotate and spread out radially in the reactor. The finely divided coal in contact with the arc area is activated and reacts with the hydrogen plasma to form acetylene at a temperature gradient of 8000–15,000 K. Acetylene concentrations in the off gas of up to 16% (yield 33 wt%) and energy consumption of 9.5 kWh/kg have been claimed. About 67% of the coal is consumed, and the remainder is converted to char, which can be used as fuel for the process.

The use of AVCO coal-based technology for the large-scale production of acetylene will depend on future oil reserves and prices. At present, it does not seem likely that coal will be a cheaper raw material than oil in the United States for this purpose. Also, the calcium carbide process, a competitive coal-based process for acetylene, has worldwide capacity with some of it in use. Its technology and economics are well understood and have been published, so that increased acetylene production can be implemented if needed without excessive development and capital costs. An improved electric furnace process for the production of 300 million pounds of acetylene per year has been reported by SRI International, together with detailed economics and process diagrams.

IX. CHEMISTRY OF SPECIALTY PRODUCTS

The chemical compounds discussed in this section are either acetylenics or their derivatives. Often, acetylene is a building block in the synthesis of a more complex molecule and is not recognizable as such in the final compound. Products such as vitamins A and E examplify this. Specialty chemicals derived from acetylene have varied and selective uses, and their market size can vary from thousands of pounds to millions of pounds per year.

A. Ethynylation Reaction

The introduction of an acetylenic group (—CH≡CH or —C≡C—) was called ethynylation in Reppe days, and the term has persisted. The reaction on a commercial scale is applied to the reaction of acetylene with aldehydes and ketones to yield acetylenic alcohols (alkynols). The ethynylation of carbonyl compounds occurs via the following routes:

1. Reppe system: Reaction of acetylene with formaldehyde in the presence of a copper acetylide catalyst under pressure
2. Lithium or sodium acetylide–liquid ammonia (atmospheric pressure)
3. Potassium hydroxide–acetylene–organic solvents (atmospheric to low pressure, 5 psig)
4. Catalytic potassium (sodium) hydroxide–acetylene–liquid ammonia (pressure, 50–400 psig)
5. Catalytic quaternary ammonium hydroxide catalyst, ethynylation of acetone under pressure (1000 psig hydrostatic)
6. Liquid acetylene as solvent and reactant (pressure 300–650 psig)
7. Grignard route: reaction of carbonyl compounds with acetylenic Grignard compounds (R—C≡C—MgBr) (atmospheric pressure)

These ethynylation methods are discussed in greater detail in Sections IX.B–IX.I.

B. Reppe Process for 2-Butyne-1,4-Diol and Propargyl Alcohol (1-Propyn-3-ol)

Cuprous acetylide deposited on silica-based carriers or supports has proved to be a unique catalyst for the reaction of aqueous formaldehyde with acetylene under pressure

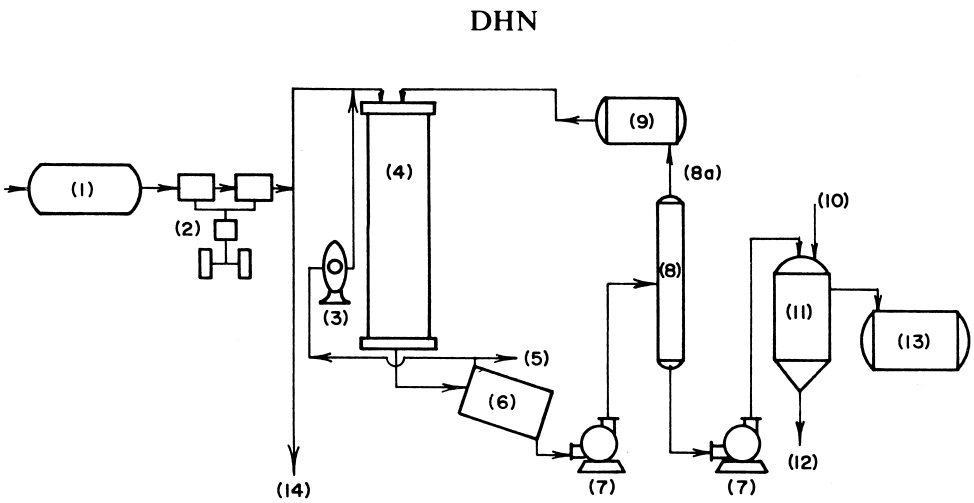


FIGURE 4 Butynediol–propargyl alcohol process. (1) Acetylene storage; (2) compressor; (3) recycle compressor; (4) pressure reactor; (5) vent; (6) separator; (7) pumps; (8) stripping still; (8a) recycle formaldehyde; (9) formaldehyde storage; (10) NaOH feed; (11) settling tank; (12) waste effluent; (13) butynediol storage; (14) recycle C_2H_2 stream for vinylpyrrolidone production. [Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 105, courtesy of Dekker, New York.]

at 90–130°C. The products resulting from this reaction, butynediol and propargyl alcohol, are formed in approximately 9:1 ratio, with butynediol the predominant product. The reaction proceeds readily (exothermic) in total conversions and selectivity of well over 90%.

$$\mathrm{HCHO} + \mathrm{C_2H_2} \xrightarrow{\mathrm{CuC_2H}} \underset{\text{Propargyl alcohol}}{\mathrm{H{-}C{\equiv}C{-}CH_2OH}}$$

$$+ \mathrm{HO{-}CH_2{-}C{\equiv}C{-}CH_2{-}OH}$$
Butynediol

Higher aldehydes or reactive methyl ketones fail to react in the Reppe system. Even a reactive aldehyde such as acetaldehyde reacts very slowly and gives only an 18% conversion in 22 hr at 125°C. In contrast, formaldehyde reacts exothermically and is the basis for the continuous operation in current Reppe plants. Flow diagrams of the butynediol–propargyl alcohol process and the continuous process for the production of all Reppe products are shown in Figs. 4 and 5, respectively. The commercial importance of these acetylene-based products and their estimated production volumes are discussed in Sections VI.A–VI.D.

C. Sodium and Lithium Acetylides in Liquid Ammonia

Although alkali metal acetylides can be prepared in polar organic solvents by the reaction of acetylene with alkali metal dispersions, the preparation and synthetic use of either sodium or lithium acetylides in liquid ammonia has proved to be the preferred route both in the laboratory and in commercial practice. Below is shown the formation of sodium acetylide in liquid ammonia and its reaction with a vitamin E intermediate, geranylacetone (GA), to yield the important terpene alkynol, dehydronerolidol (DHN):

$$1.\ \mathrm{Na} + \text{Liquid}\,\mathrm{NH_3} \xrightarrow[-33^\circ\mathrm{C}]{\mathrm{Fe(NO_3)_3}} \mathrm{NaNH_2} + \tfrac{1}{2}\mathrm{H_2} \qquad (3)$$

$$2.\ \mathrm{NaNH_2} + \mathrm{C_2H_2} \xrightarrow{\mathrm{NH_3}} \mathrm{NaC_2H} + \mathrm{NH_3}$$

$$3.\ (\mathrm{CH_3})_2{-}\mathrm{C}{=}\mathrm{CH}{-}(\mathrm{CH_2})_2{-}\underset{\text{GA}}{\overset{\mathrm{CH_3}}{\overset{|}{\mathrm{C}}}}{=}\mathrm{CH}{-}(\mathrm{CH_2})_2$$
$$-\mathrm{COCH_3} + \mathrm{C_2H(NH_3)}$$
$$\downarrow \begin{matrix}1.\mathrm{NaC_2H}\\ 2.\mathrm{H_2O{-}NH_4Cl}\end{matrix} \qquad (4)$$

$$(\mathrm{CH_3})_2{-}\mathrm{C}{=}\mathrm{CH}{-}(\mathrm{CH_2})_2{-}\overset{\mathrm{CH_3}}{\overset{|}{\mathrm{C}}}{=}\mathrm{CH}{-}(\mathrm{CH_2})_2{-}\underset{\mathrm{OH}}{\overset{\mathrm{CH_3}}{\overset{|}{\underset{|}{\mathrm{C}}}}}{-}\mathrm{C}{\equiv}\mathrm{CH}$$
DHN
(5)

This technology is used by Hoffmann–La Roche to prepare intermediates for the manufacture of vitamins A and E. The process has been in operation since the early 1950s. The use of lithium acetylide is preferred for more sensitive or less reactive carbonyl compounds. Reactions (3)–(5) are run consecutively in liquid ammonia without isolation. The formation of the alkali metal acetylide (sodium,

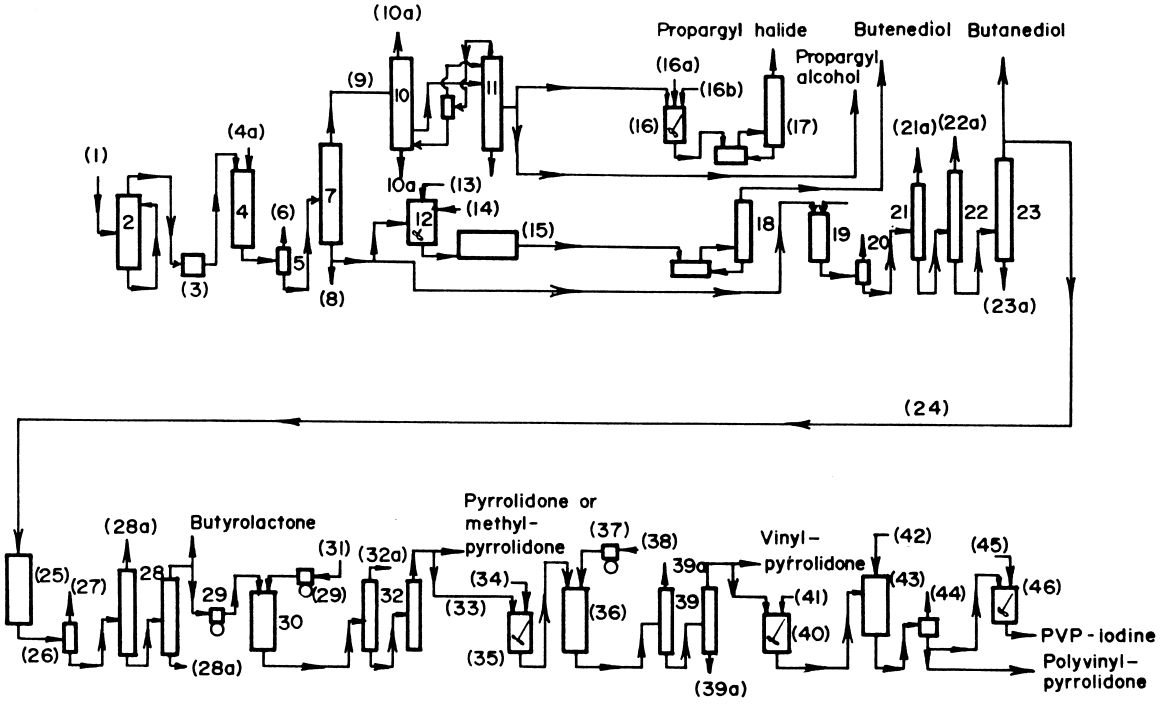


FIGURE 5 Continuous process for the production of Reppe chemicals (GAF). (1) Crude acetylene; (2) scrubber; (3) compressor; (4) ethynylation reactor; (4a) formaldehyde; (5) separator; (6) vent; (7) still; (8) butynediol; (9) crude propargyl alcohol; (10) propargyl alcohol still; (10a) to waste; (11) propargyl alcohol still; (12) batch hydrogenation; (13) hydrogen feed; (14) catalyst feed; (15) catalyst filter; (16) halogenation reactor; (16a) catalyst; (16b) halogen; (17) propargyl halide still; (18) butenediol batch still; (19) butanediol hydrogenator; (20) vent; (21) butanediol still; (21a) to butanol recovery; (22) butanediol still; (22a) waste; (23) final butanediol still; (23a) waste; (24) to dehydrogenation reactor; (25) dehydrogenation reactor; (26) separator; (27) to hydrogen recovery; (28) butyrolactone stills; (28a) waste; (29) pump; (30) pyrrolidone product reactor; (31) liquid NH_3 or methylamine feed; (32) product stills; (32a) waste; (33) pyrrolidone feed; (34) catalyst feed; (35) catalyst preparation; (36) vinylation reactor; (37) compressor; (38) purified acetylene; (39) vinylpyrrolidone stills; (39a) waste; (40) polymerizer; (41) catalyst and distilled water; (42) purified air; (43) spray dryer; (44) vent; (45) iodine feed; (46) reactor. [Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," p. 106, courtesy of Dekker, New York.]

lithium) is rapid and quantitative, and the formation of the alkynol averages more than 90%.

D. Potassium Hydroxide–Acetylene–Organic Solvent Route

Favorsky, the Russian chemist, discovered that potassium hydroxide ground in solvents such as ethers, acetals, and amines was an effective medium for reacting acetylene with ketones and certain aldhydes to yield acetylenic alcohols and glycols. This reaction system yields secondary and tertiary acetylenic hydroxy compounds and is particularly useful for the tertiary series. This technology is now practiced by Air Products and Chemicals via a continuous process whereby both acetylenic alcohols and diols are produced in organic media.

The ethynylation reaction, in general, can be controlled via acetylene concentration, reaction temperature, and the ratio of base to ketone and acetylene. Alkynol formation is favored at lower temperatures, generally below 10°C, while alkynediol formation is favored at 30–50°C. What follows is the formation of methylbutynol and dimethylhexynediol from acetone and acetylene:

$$(CH_3)_2{-}CO + C_2H_2 \xrightarrow[2.\ H_2O]{1.\ KOH-solvent}$$

$$(CH_3)_2{-}\underset{\displaystyle OH}{\underset{|}{C}}{-}C{\equiv}CH + (CH_3)_2{-}\underset{\displaystyle OH}{\underset{|}{C}}{-}C{\equiv}C{-}\underset{\displaystyle OH}{\underset{|}{C}}{-}(CH_3)_2$$

Acetylenic alcohols and diols produced commercially by this process are shown in the tabulation below.

Products with the (S-) designation are Surfynol surfactants extensively used as nonfoaming wetting and dispersing agents. The applications for these products are discussed in Section VII.C.

E. Ethynylation in Liquid Ammonia Using Catalytic Amounts of KOH or Sodium Acetylide

Tedeschi showed that catalytic amounts of either NaOH or KOH in liquid ammonia–acetylene medium could affect the ethynylation of aldehydes and ketones in high catalytic efficiency. The reaction was carried out at 20–40°C at acetylene–ammonia pressures of 150–300 psig. The mole ratio of reactants for methylbutynol formation averages 6–18 mol acetone, 18–24 mol acetylene, 0.5–1.5 mol KOH, and 18–24 mol ammonia. The conversion to methylbutynol based on acetone averages 75–100%, with a catalytic yield based on KOH averaging 380–900%. Longer chain aldehydes and ketones react in the ammonia system with equal ease. The ethynylation of methylheptenone to DHL proceeds catalytically in more than 90% conversion. When stoichiometric to excess amounts of acetylene are used in the ammonia system, acetylenic diol formation is quite low (0.5–1.0%).

The reaction proceeds via the formation of an alkynol–KOH complex. the isolated crystalline complex can be substituted for KOH in the ethynylation reaction with equal results. However, lithium hydroxide, unlike lithium acetylide in liquid ammonia, fails to yield any acetylenic hydroxy compound. The mechanism of catalytic ethynylation in liquid ammonia and aprotic solvents based on key intermediates has been published by Tedeschi. Alkali metal acetylides can be used as catalytic species in place of KOH. Although acetylenic glycols can be formed in liquid ammonia using KOH, the reaction is not catalytic, requiring stoichiometric to excess amounts of base. The catalytic liquid ammonia–KOH process has great potential for economically producing secondary and tertiary alkynols. Long-chain terpenoid alkynols used as intermediates in the synthesis of vitamins A and E can be made in high conversions and catalytic efficiency in the ammonia system. Secondary alkynols, such as 1-hexyn-3-ol and 4-ethyl-1-octyn-3-ol used commercially as corrosion inhibitors, are also readily formed. Methylbutynol, made from acetone and acetylene, is important commercially as a starting point for vitamins A and E, as a stabilizer for chlorinated solvents, as an intermediate for the herbicide Kerb, and for the large-scale production of isoprene. In the last-named use it is made catalytically in liquid ammonia, as described in Section IX.F.

Carbonyl compound	Alkynol	Alkynediol
Acetone	Methylbutynol	Dimethylhexynediol
Methyl ethyl ketone	Methylpentynol	Dimethyloctynediol (S-82)
Methyl isobutyl ketone	Dimethylhexynol (S-61)	Tetramethyldecynediol (S-104)
Butyraldehyde	Hexynol	
2-Ethylhexaldehyde	Ethyloctynol	

F. Isoprene from Acetylene and Acetone

Anic (an affiliate of Ente Nazionale Idocarburi) at Ravenna, Italy, has been operating a large production facility for the manufacture of isoprene, which is used to make *cis*-polyisoprene. The route developed by Anic involves the catalytic formation of methylbutynol (MB), semihydrogenation to methylbutenol (MBe), and dehydration of the latter to isoprene:

$$1.\quad (CH_3)_2{-}CO + C_2H_2 \xrightarrow[50\%\ KOH]{NH_3} (CH_3)_2{-}\underset{\underset{MB}{OH}}{C}{-}C{\equiv}CH \qquad (6)$$

$$2.\quad MB + H_2 \xrightarrow{Pd\ catalyst} (CH_3)_2{-}\underset{\underset{MBe}{OH}}{C}{-}CH{=}CH_2 \qquad (7)$$

$$3.\quad MBe \xrightarrow{-H_2O} H_2C{=}C(CH_3){-}CH{=}CH_2 \qquad (8)$$

Methylbutynol is formed continuously in liquid ammonia under pressure using 50% aqueous KOH as catalyst. The conversions to both MB and MBe are close to 100%, and the overall yield to isoprene is well over 90%. The process delivers high-purity isoprene, readily purified for use in the isotactic polymerization of isoprene to *cis*-polyisoprene. A process flow diagram of the Anic methylbutynol–isoprene process (SNAM Progetti) is shown in Fig. 6. The plant originally had a capacity of 66 million pounds of isoprene per year but is believed to have been expanded.

G. Grignard Route

The use of Grignard reagents to form acetylenic products is a relatively expensive route used primarily in the pharmaceutical industry to produce drugs or vitamin intermediates. The acetylenic Grignard reagent is readily formed in 100% yield by reacting ethylmagnesium bromide with a terminal acetylenic compound in THF media:

$$H_5C_2{-}MgBr + R{-}C{\equiv}CH \rightarrow R{-}G{\equiv}C{-}MgBr + C_2H_6$$

The acetylenic Grignard is used *in situ* and is best employed with sensitive aldehydes or ketones, which are affected adversely by basic reagents. An intermediate step in the manufacture of vitamin A involves the formation of the 1,5-acetylenic di-Grignard, which is then reacted

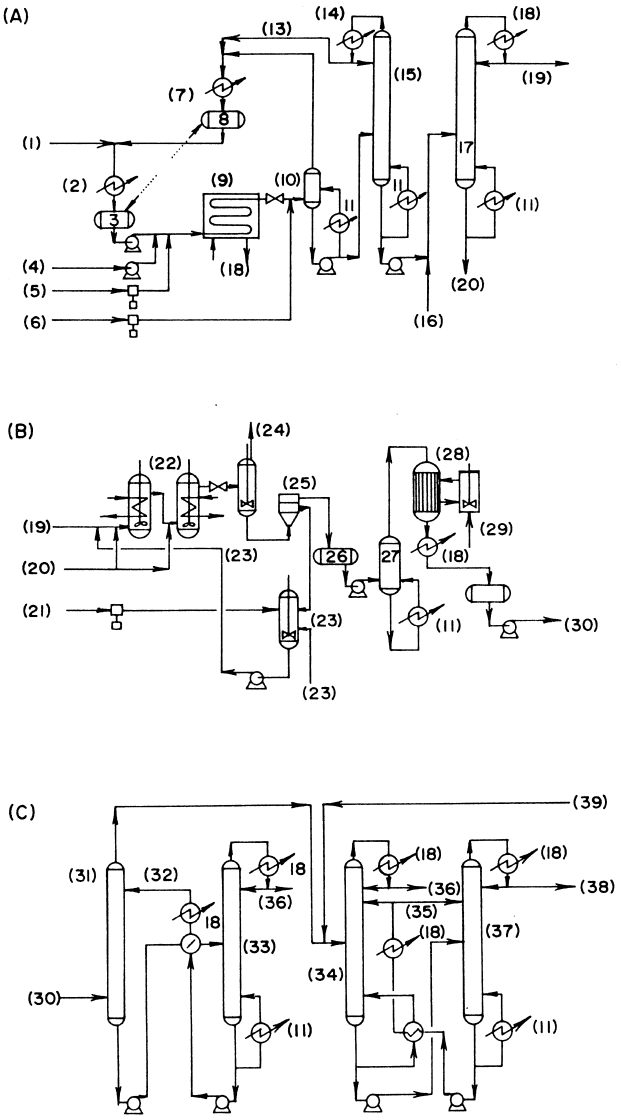


FIGURE 6 Anic (SNAM Progetti) process: isoprene from acetylene and acetone. (A) Methylbutynol; (B) methylbutenol and isoprene; (C) isoprene purification. (1) Acetylene feed; (2) refrigerant; (3) ammonia condenser; (4) acetone feed; (5) KOH catalyst feed; (6) stopper; (7) refrigerant; (8) acetylene condenser; (9) ethynylation reactor; (10) flash tank; (11) steam; (12) NH_3 + unreacted C_2H_2; (13) unreacted acetone; (14) refrigerant; (15) acetone recovery column; (16) water; (17) heavy ends column; (18) cooling water; (19) aqueous methylbutynol to hydrogenation; (20) hydrogen feed; (21) inhibitor feed; (22) hydrogenation reactors; (23) fresh catalyst and catalyst feed; (24) unreacted hydrogen; (25) catalyst separator; (26) methylbutenol (MBe) receiver; (27) MBe evaporator; (28) dehydratior reactor; (29) fuel gas; (30) raw isoprene to purification; (31) washing tower; (32) water to washing tower; (33) water distillation; (34) extractive distillation column; (35) extraction solvent; (36) purge line; (37) isoprene stripper; (38) pure isoprene; (39) reject isoprene from polymerization plant. (Reprinted from Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," pp. 164–165, courtesy of Dekker, New York.)

with a β-ionylidene (C_{14}) aldehyde to yield an acetylenic dihydroxy vitamin A precursor, as illustrated below. In general, the yields are high and the conditions mild for the Grignard route. When acetylene is reacted with ethylmagnesium bromide, both hydrogen atoms are replaced and the resulting acetylenedimagnesium bromide is exclusively formed. This di-Grignard can be used to form sensitive acetylenic diols (see Scheme 1).

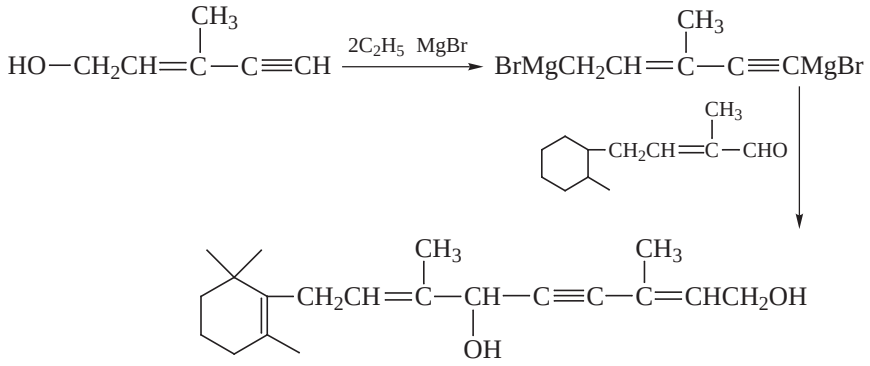
H. Use of Basic Quaternary AmmonHium Hydroxide Resins

Considerable effort has been expended on the development of basic resins as replacements for KOH or NaOH in the catalytic ethynylation of carbonyl compounds. The objective in the past has been to develop a fixed-bed continuous process in which neutralizaion of the reaction mixture is not required and product isolation is much simplified. This type of process has been extensively studied in the catalytic ethynylation of acetone to form methylbutynol.

The quaternary ammonium hydroxide resin is prepared from the chloride form of the resin by elution with sodium hydroxide solution, followed by washing with water, then methanol washing, and finally vacuum drying. A serious problem associated with these resins is their degradation and loss of activity on continued use. Furthermore, aldehydes react very poorly and give almost entirely aldol byproducts. Ketones higher than methyl ethyl ketone react so slowly that the process is not economical. The use of liquid ammonia has been employed to activate the resin and reactants. However, due to the limited application of this process, it is doubtful whether this technology is widely used or whether it is used at all.

I. Use of Liquid Acetylene as Solvent and Reactant

Liquid acetylene employed below its critical pressure (650 psi) and temperature (36°C) was used at the bench-scale level, in batch reactors, to prepare such products as alkali metal acetylides (lithium, sodium, potassium), methylbutynol, aminobutynes, and transition metal complexes. The work was carried out under 600 psig acetylene pressure in a special reactor system and pressure cubicle. The research showed that liquid acetylene could be handled safely in a variety of syntheses, with no exothermic decompositions or explosions. Pure, white alkali metal acetylides were easily formed at 10–25°C, and the formation of methylbutynol took place in high conversions (80–85%) using only catalytic amounts of KOH and ammonia. This work by Tedeschi and Moore was never scaled up to commercial practice.



SCHEME 1

X. VINYL ETHERS

The reaction of alcohols with acetylene to form vinyl ethers, commonly termed vinylation, takes place readily at 150–180°C using basic catalysts such as alkali metal hydroxides or alkoxides:

$$ROH + C_2H_2 \rightarrow RO{-}CH{=}CH_2$$

The vinylation process can be carried out at both atmospheric pressure and under elevated pressures with an acetylen–nitrogen atmosphere. The formation of methyl vinyl ether from methanol and acetylene can be carried out continuously at atmospheric pressure and 180°C. The GAF Corporation has produced a variety of vinyl ethers derived from methoanol, ethanol, *n*-butanol, isobutanol, and decanol. The vinyl ethers find their most important use as the copolymers derived from methyl vinyl ether and maleic anhydride, known as the Gantrez AN series. Applications of these products and other polyvinyl ethers are summarized in Section VI.A.

XI. FLAVOR AND FRAGRANCE COMPOUNDS AND VITAMINS A AND E

The reaction flow sheets in Figs. 7–13 summarize the rather lengthy chemistry and technology developed since 1955 for the production of polyterpene compounds. This chemistry is characterized by change and is highly proprietary. Prominent companies in this field are Hoffmann–La Roche and BASF. The choice of a given synthetic route depends on cost and availability of raw materials in addition to manufacturing sites.

The synthesis sequence pioneered by Roche and used for more than 30 years involves (1) ethynylation, (2) semihydrogenation, (3) reaction with diketene or methylisopropenyl ether, and (4) allylic isomerization with loss of carbon dioxide. This four-step sequence is repeated a number of times in the synthesis of both vitamins A (*trans*-retinol) and E (*dl*-α-tocopherol). Some of the intermediate products such as methylheptenone, linalool, pseudoionone, β-ionone, citral, geranylacetone, and nerolidol, are valuable flavor–perfumery chemicals. This circumstance increases the overall versatility and profitability of the process by providing diversification.

The reaction sequence in Fig. 7 illustrates the following transformations:

$$\text{Acetone} + \text{acetylene} \rightarrow \text{methylbutynol}$$
$$\rightarrow \text{methylbutenol} \rightarrow \text{methylheptenone}$$
$$\rightarrow \text{dehydrolinalool} \rightarrow \text{linalool}$$

This route and related routes are high-yield, high-selectivity steps of basic importance to the final economics of the processes for vitamins A and E. The second sequence, shown in Fig. 8, illustrates two processes for the production of citral from the alkynol DHL. The DHL–acetic anhydride route involves earlier technology, which may have been replaced by the direct isomerization of DHL, the second process. Large amounts of citral are also produced by SCM-Glidden via the oxidative dehydrogenation of geraniol.

The flow sheet shown in Fig. 9 illustrates the following process sequences:

$$\text{DHL or citral} \rightarrow \text{pseudoionone} \rightarrow \alpha\text{- and } \beta\text{-ionones;}$$

$$\text{linalool} \rightarrow \text{geranylacetone} \rightarrow \text{dehydronerolidol}$$
$$\rightarrow \text{nerolidol}$$

Citral and DHL compete with one another in the preparation of pseudoionone, the precursor of the ionones. Lemongrass oil and myrcene (from turpentine) are natural sources of citral. β-Ionone is a very important

Product sequence:

acetone + acetylene ⟶ methylbutynol (MB) ⟶ methylbutenol (MBe) ⟶ methylheptenone (MHe)

methylheptenone (MHe) ⟶ dehydrolinalool (DHL) ⟵ linalool

Technology:

$$CH_3COCH_3 + C_2H_2 \xrightarrow{(1)} (CH_3)_2C(OH)C{\equiv}CH \xrightarrow[H_2]{(2)} (CH_3)_2C(OH)CH{=}CH_2 \xrightarrow[DK]{(3)} (CH_3)_2C(OCOCH_2COCH_3)CH{=}CH_2$$

(MB, 90–95%) (MBe, 97–99%)

$\xrightarrow[(4)]{\Delta,\ 160\text{–}180^\circ C}$

$$(CH_3)_2C{=}CHCH_2CH_2C(CH_3)(OH)CH{=}CH_2$$

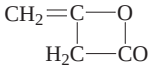
(linalool, 97–99%)

$$(CH_3)_2C{=}CHCH_2CH_2C(CH_3)(OH)C{\equiv}CH \xrightarrow[(5)]{C_2H_2} (CH_3)_2C{=}CHCH_2CH_2COCH_3 + CO_2$$

$\xleftarrow[(6)]{H_2}$

(DHL, 90–95%) (MHe, 80–90%)

FIGURE 7 Linalool and dehydrolinalool from acetylene–acetone (Roche, BASF). Process conditions (average process yields under formulas): (1,5) catalytic ethynylation in polar solvents using KOH or sodium under pressure ($t = 30$–50°C); (2,6) partial hydrogenation using a palladium catalyst modified ($t = 25$–45°C); (3) reaction of methylbutenol with diketene (DK)



($t = 20$–40°C): catalyst, tertiary amine; MBe, acetocetate ester not isolated [see (4)]; (4) allylic rearrangement and cleavage of product (3) using aluminum isopropoxide under fractionation column.

From dehydrolinalool (DHL): (Roche technology)

$$(CH_3)_2C{=}CH(CH_2)_2C(CH_3)(OH)C{\equiv}CH + (CH_3CO)_2O \longrightarrow (CH_3)_2C{=}CH(CH_2)_2C(CH_3)(OCOCH_3)C{\equiv}CH + CH_3CO_2H$$

(DHL) (not isolated)

$\xrightarrow[Ag\ Salt]{(3)\ \Delta}$

$$(CH_3)_2C{=}CH(CH_2)_2C(CH_3){=}CHCHO \xleftarrow{H_2O} (CH_3)_2C{=}C(CH_2)_2C(CH_3){=}C(OCOCH_3)_2 + (CH_3)_2C{=}CH(CH_2)_2C(CH_3){=}C{=}CHOCOCH_3$$

(citral, overall yield ~82%) (allenic acetates, not isolated)

Isomerization of DHL: DHL ⟶ citral

FIGURE 8 Processes for the production of citral. Roche: Catalysts, siloxyvanadium moieties; excess tris(trimethylsiloxy) vanadium oxide; conditions, liquid phase, 2–20 hr at 125–140°C; conversions 72–77%; yield 96%. Rhone Poulenc (Rhodia): Catalysts, vanadium alkoxides, excess cyclohexyl orthovanadate; conditions, liquid phase, 15 min to 4 hr at 125–160°C; conversions 37–40%; yield 77–82%; Oxidative Dehydrogenation of Geraniol (SCM-Glidden).

$$(CH_3)_2C(CH_3)CH(CH_2)2C{=}CHCH_2OH \xrightarrow[O]{-H} Citral + H_2O$$

Possibly vapor-phase process; copper-type catalyst; estimated yield 70–80%.

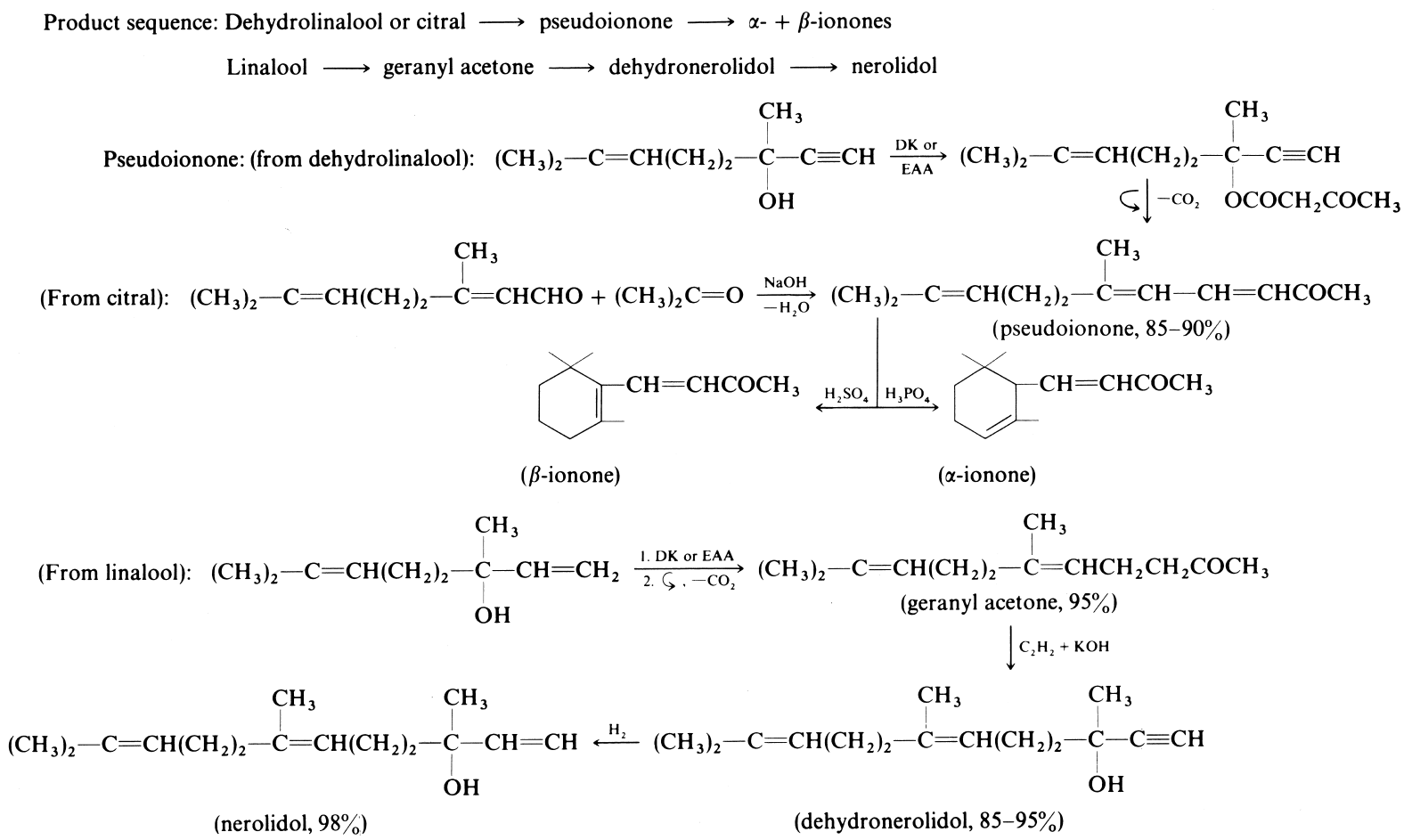


FIGURE 9 Higher terpene flavor and fragrance compounds derived from citral, dehydrolinalool, and linalool. DK, Diketene; EAA, ethyl acetoacetate.

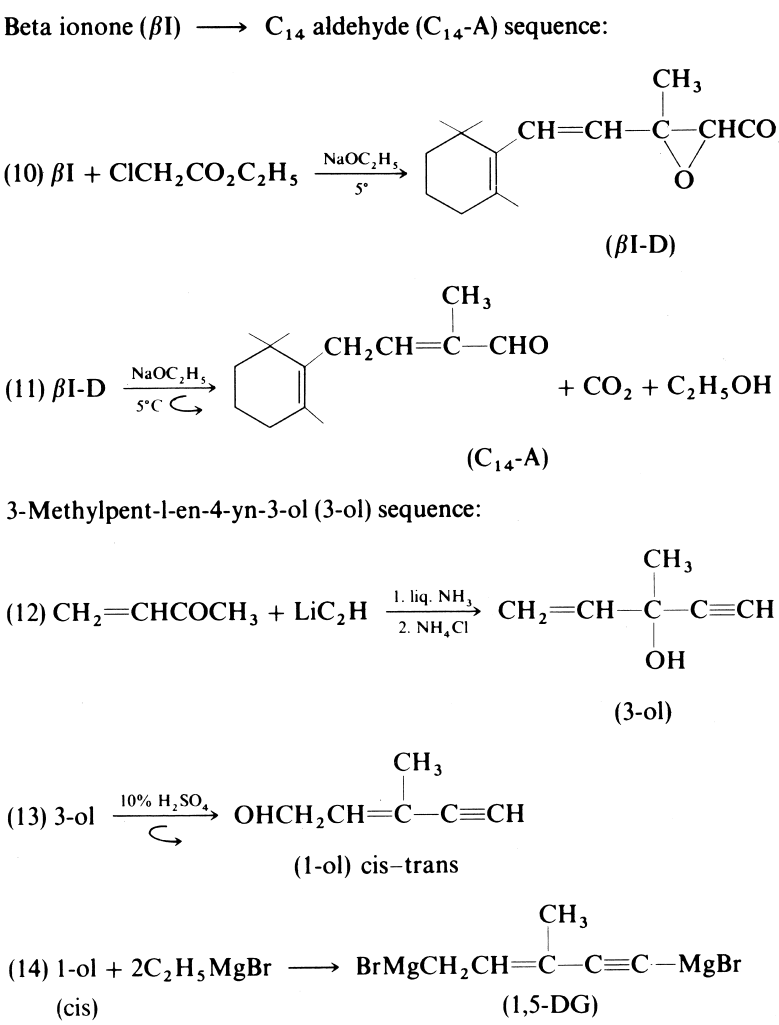


FIGURE 10 Vitamin A production. Reaction flow sheet (steps 10–14).

intermediate, since it contains the correct isomeric ring system for vitamin A. Linalool, made from DHL via partial hydrogenation, is a key intermediate in the synthesis of vitamin E (*dl*-α-tocopherol) via the above compounds, leading to isophytol, the precursor of *dl*-α-tocopherol.

Vitamin A production is shown by the reaction flow sheets in Figs. 10 and 11: β-ionone (βI) → C_{14} aldehyde, and C_{14} aldehyde → vitamin A, respectively. β-Ionone reacted with chloroethyl acetate via the Darzens reaction is converted to the intermediate epoxy ester, which on rearrangement yield C_{14} aldehyde.

The di-Grignard reagent (Fig. 10) of 3-methyl-pent-2-en-4-yn-1-ol (*cis*-1-ol) on reaction with C_{14} aldehyde (Fig. 11) yields the C_{20} eneynediol (C_{20} yn-D), which contains the correct ring system and carbon chain of vitamin A. The important *cis*-1-ol compound is derived from methyl vinly ketone and lithium acetylide, followed by isomerization. The use of lithium acetylide is preferred since sodium acetylide gives considerably lower yields. The remaining four steps of the process [(16)–(18)] are concerned with the synthesis of the all-trans vitamin A (retinol).

The flow sheet in Fig. 12 shows the 12-step process leading to the C_{20} polyterpene alcohol isophytol (IP). The formation of vitamin E is shown in Fig. 13, as are syntheses for the second component, trimethylbenzoquinone (TMHQ). The condensation of IP and TMHQ yields *dl*-α-tocopherol (vitamin E).

Vitamin A is also manufactured by Wittig technology, starting with β-ionone. This route was pioneered by G. Wittig and H. Pommer at BASF. A general synthesis step involves the condensation of terpene alcohols, such as vinyl-β-ionol, with triphenylphosphine in the presence of HCl to yield quaternary phosphonium salts. These

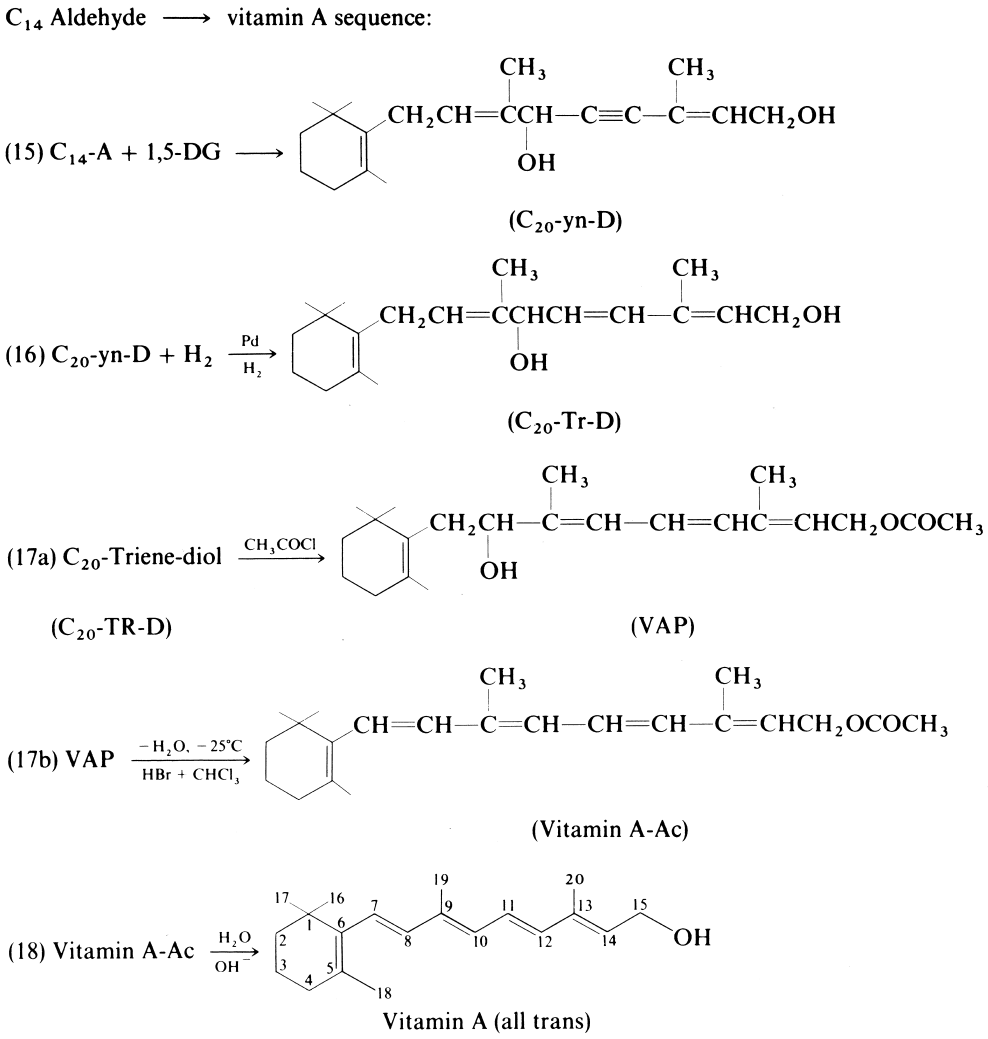


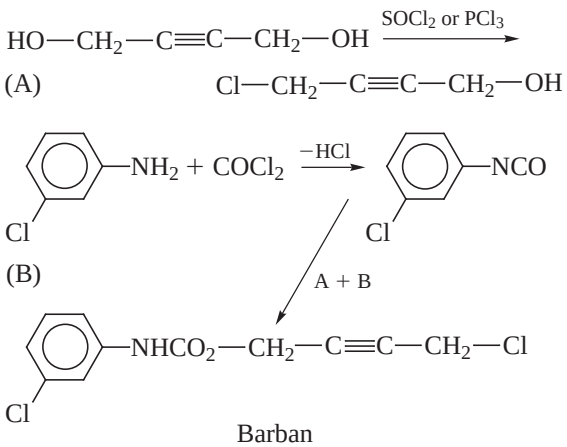
FIGURE 11 Vitamin A production. Reaction flow sheet (steps 15–18): C_{14} aldehyde → vitamin A sequence.

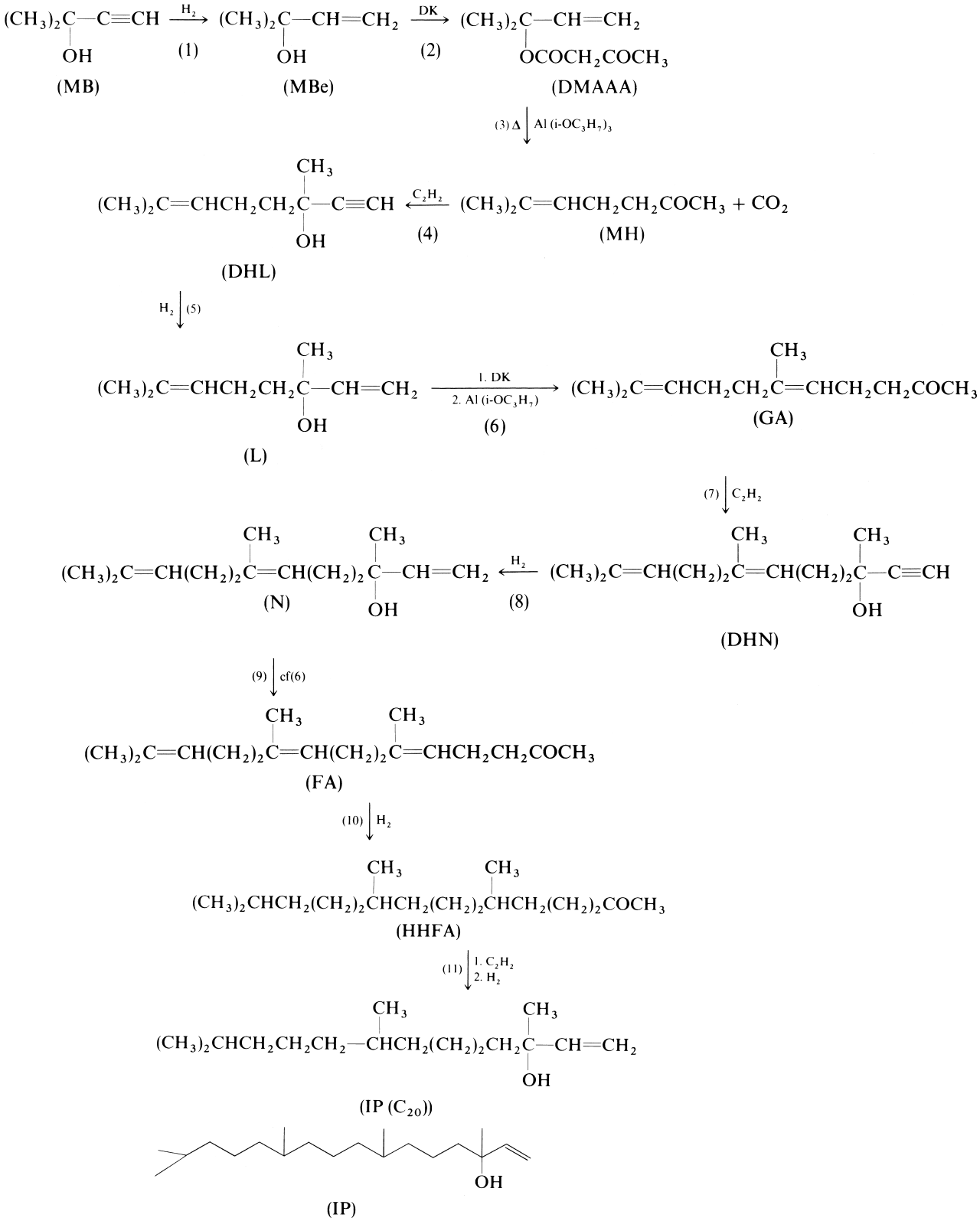
salts in the presence of base undergo rapid elimination to phosphoranes (ylides), which react readily with carbonyl compounds to form olefins to high yield. By a repetition of this sequence the carbon chain is built up to that of retinol.

XII. ACETYLENIC PESTICIDES

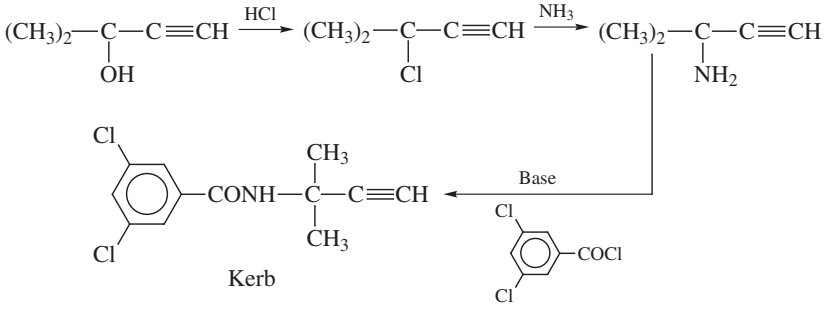
Acetylenic compounds that have attained commercial status as pesticides are Barban (carbyne), Kerb (pronamide), and Omite (propargite). The last-named compound is a miticide, while the other two are herbicides. Their uses are outlined in Section VII.E. Their synthesis is summarized as follows:

A. Barban

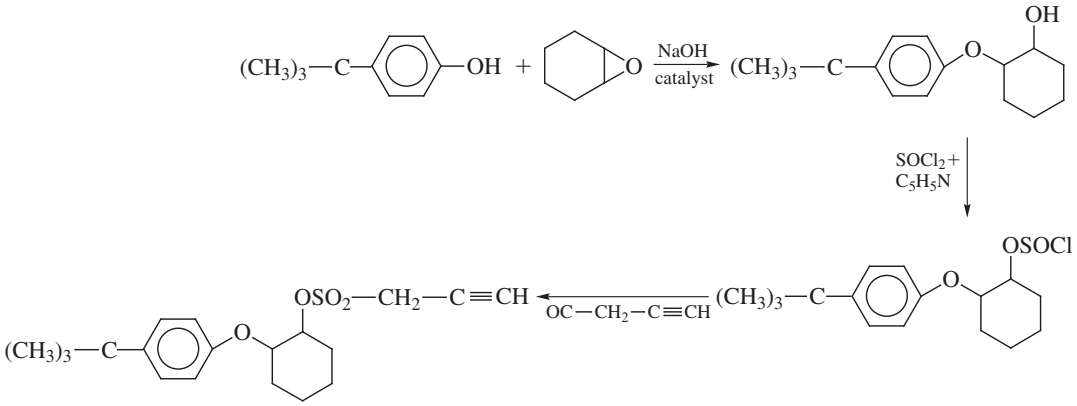



(CH3)2C—C≡CH
OH
(MB)
H2
(1)
(CH3)2C—CH=CH2
OH
(MBe)
DK
(2)
(CH3)2C—CH=CH2
OCOCH2COCH3
(DMAAA)
(3) Δ
Al (i-OC3H7)3
CH3
(CH3)2C=CHCH2CH2C—C≡CH
OH
(DHL)
C2H2
(4)
(CH3)2C=CHCH2CH2COCH3 + CO2
(MH)
H2
(5)
CH3
(CH3)2C=CHCH2CH2C—CH=CH2
OH
(L)
1. DK
2. Al (i-OC3H7)
(6)
CH3
(CH3)2C=CHCH2CH2C=CHCH2CH2COCH3
(GA)
(7)
C2H2
CH3
CH3
(CH3)2C=CH(CH2)2C=CH(CH2)2C—CH=CH2
OH
(N)
H2
(8)
CH3
CH3
(CH3)2C=CH(CH2)2C=CH(CH2)2C—C≡CH
OH
(DHN)
(9)
cf(6)
CH3
CH3
(CH3)2C=CH(CH2)2C=CH(CH2)2C=CHCH2CH2COCH3
(FA)
(10)
H2
CH3
CH3
(CH3)2CHCH2(CH2)2CHCH2(CH2)2CHCH2(CH2)2COCH3
(HHFA)
(11)
1. C2H2
2. H2
CH3
CH3
(CH3)2CHCH2CH2CH2—CHCH2(CH2)2CH2C—CH=CH2
OH
(IP (C20))
OH
(IP)

B. Kerb



C. Omite



XIII. ACETYLENIC REACTIONS WITH RESEARCH AND COMMERCIAL POTENTIAL

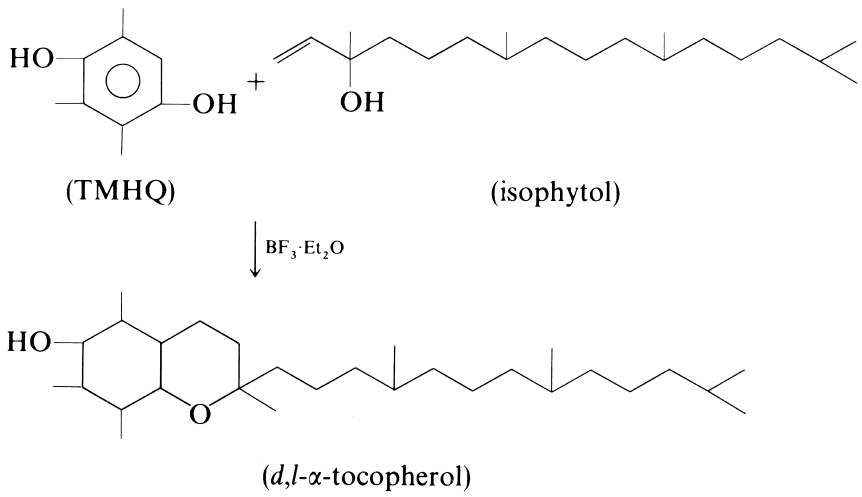
The products and reactions described in this section have not been extensively developed on a commercial scale but are believed to have potential in fine-chemicals applications.

A. Cyclic Carbonates from Alkynol–Alkynediol–Carbon Dioxide Reactions

An interesting reaction discovered by Dimroth and Pasedach at BASF involves the reaction of tertiary acetylenic alchols and diols with carbon dioxide under pressure to form substituted cyclic carbonates (5-methylene-4,4-di-alkyldioxolan-2-ones). The reaction is catalyzed by the combination of a cuprous salt and a tertiary

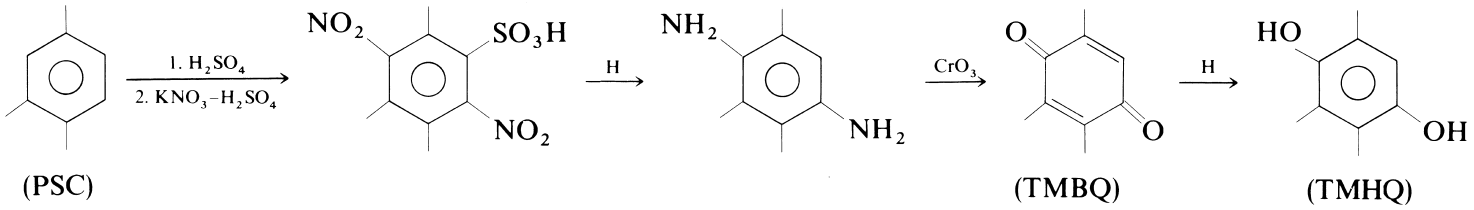
FIGURE 12 Vitamin E: reaction sequence leading to isophytol.

Abbreviation	Common name	Process step and reaction
MB	Methylbutynol	(0) Ethynylation of acetone
MBe	Methylbutenol	(1) Semihydrogenation (Pd)
DMAAA	Dimethylallyl acetoacetate	(2) Diketene (DK), ketenization
MH	Methylheptenone	(3) Allylic isomerization (Carrol)
DHL	Dehydrolinalool	(4) Ethynylation of MH
L	Linalool	(5) Same as (1)
GA	Geranylacetone	(6) Same as (2) and (3)
DHN	Dehydronerolidiol	(7) Same as (4)
N	Nerolidol	(8) Same as (1) and (5)
FA	Farnesylacetone	(9) Same as (2), (3), and (6)
HHFA	Hexahydrofarnesylacetone	(10) Complete hydrogenation, side chain
IP	Isophytol	(11) Same as (4) and (5)



Trimethyl hydroquinone–trimethyl benzoquinone synthesis:

From pseudocumene (PSC):



From 2,3,6- or 2,3,5-trimethyl phenol:

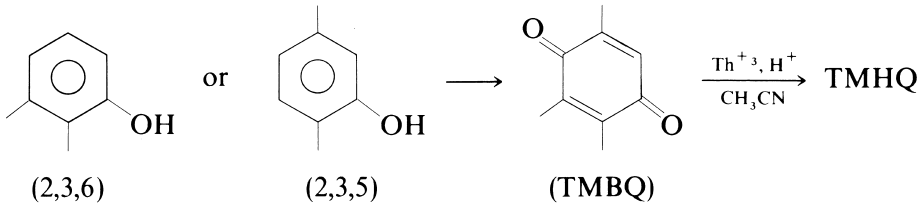
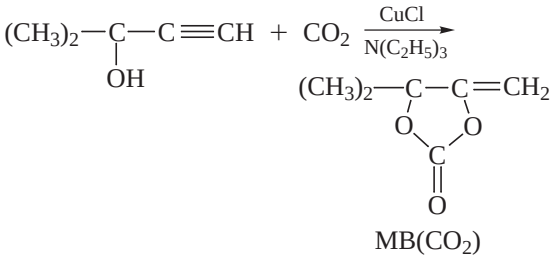


FIGURE 13 Final conversion step to vitamin E (α-tocopherol), TMHQ, trimethylhydroquinone; TMBQ, trimethylben zoquinone; PSC, pseudodocumene.

amine and is generally carried out at 30–50 atm and temperatures of 80–120°C.



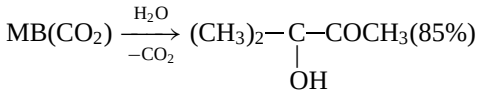
In the presence of excess, gaseous CO_2 at 80°C and 800 psig, methylbutynol reacts in 12 hr to form the cyclic carbonate MB(CO_2) in 90% yield. Liquid CO_2 has also been used successfully as both solvent and reactant below its critical temperature (35°C) with secondary and tertiary acetylenic alcohols, 1-haloalkynols, and tertiary acetylenic glycols. The mass action and solvent effect of liquid CO_2 makes it possible for these reactions to proceed readily at lower temperature and pressures, at faster reaction rates, and with fewer by-products, particularly with secondary alkynols. However, the reaction is unsuccessful with propargyl alcohol or secondary acetylenic glycols, illustrating an unusual reaction specificity. The reaction of dimethyloctadiynediol (oxidative coupling product of methylbutynol) and CO_2 leads to the formation of a bifunctional carbonate with interesting potential in polymer chemistry. Acetylenic alcohols and diols reacted successfully with liquid CO_2 at 20–30°C under pressure (500 psig) are listed in Table X. Compounds (a), (b), (c), (e), and (f) are tertiary alkynols or alkynediols, while (d) is a secondary alkynol. The reaction gives highest yields with the tertiary series.

The versatility of the cyclic carbonates is illustrated by typical reactions carried out by Pasedach and co-workers. The alkynol (MB) is reacted *in situ* to form the cyclic carbonate MB(CO_2), which then undergoes further reaction in the presence of an active hydrogen reactant.

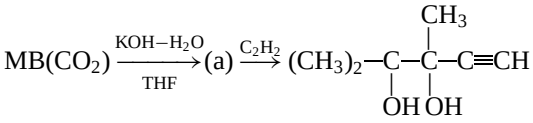
Hydroxy ketones can be formed via reaction with water, followed by loss of CO_2. This is an alternative method of hydrating a triple bond without the use of mercury salts or expensive transition metal catalysts. Reaction with alcohols in turn yields the corresponding methylketo open-chain carbonate in high yield (88–90%). Ethynyldiols result in good yield through the sequential conversion of the carbonate to the hydroxy ketone via water present in the KOH, followed by ethynylation of the carbonyl precursor. The substitution of excess primary amine for the tertiary amine cocatalyst results in good yields of disubstituted methylene oxazolidones. The reactions are carried out at temperatures of 80–150°C at pressures below 40 atm.

TABLE X Acetylenic Compounds Reacted with Liquid Carbon Dioxide

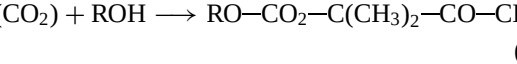
Acetylenic compound	Yield of cyclic carbonate (%)
(a) 3-Methyl-1-butyn-3-ol	90
(b) 1-Chloro-3-methyl-1-butyn-3-ol	76
(c) 1-Ethynylcyclohexanol	87
(d) 4-Methyl-1-pentyn-3-ol	66
(e) 2,5-Dimethyl-3-hexyne-2,5-diol	93
(f) 2,7-Dimethylocta-3,5-diyne-2,7-diol	76



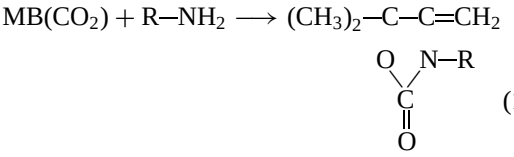
(9)



(10)



(11)



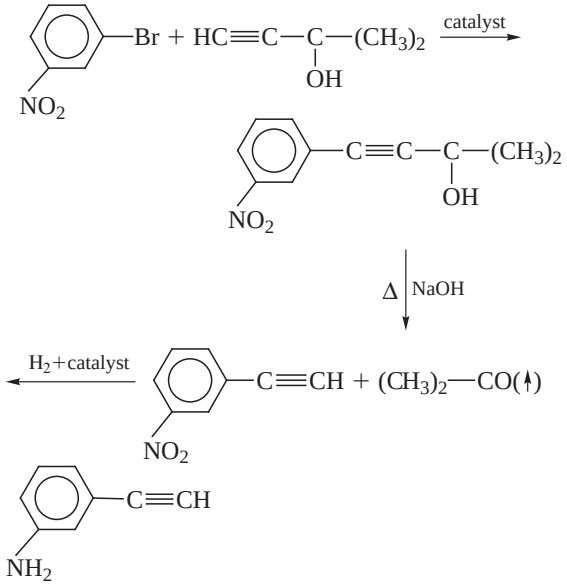
(12)

The cyclic carbonate from methylbutynol MB(CO_2) can also be copolymerized with such monomers as methyl acrylate, styrene, and acrylonitrile. The acrylate copolymer was considerably harder than the homopolymer of methyl acrylate and had a glass transition temperature of 90°C vs 9°C for polymethylacrylate. By virtue of its carbonate functionality, MB(CO_2) is an interesting reactive monomer that not only provides increased hardness and strength to the polymer matrix, but can also be used to make modified polymer structures by further reaction with the reactive ring system.

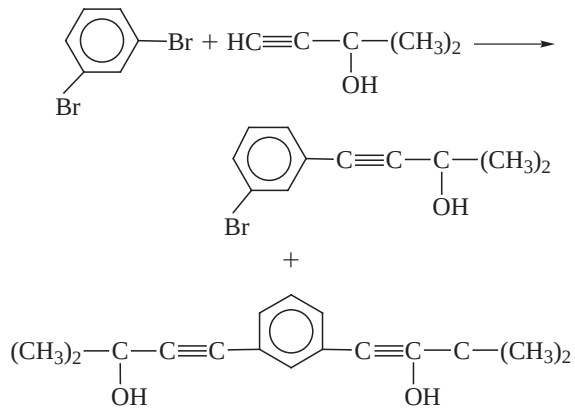
B. Arylacetylenes from Methylbutynol and Aromatic Compounds

Selwitz and co-workers at Gulf Research and Development Company have disclosed interesting technology for the production of substituted arylacetylenes, where the ethynyl group is directly attached to the aromatic nucleus. Arylacetylenes are normally quite difficult to prepare by direct substitution. The important features of this process are the reaction of an active aryl

such as *m*-nitrobromobenzene with 3-methyl-1-butyn-3-ol (MB), using a catalyst composed of palladious chloride and triphenylphosphine. The catalyst complex $[PdCl_2–P–(C_6H_5)_3]$ is further activated with a catalytic amount of cuprous iodide, used as promoter. The reaction is carried out in an amine solvent at 50–100°C. The resulting nitrophenylhydroxyacetylene is then cleaved with base to the *m*-nitrophenylacetylene and acetone. The base cleavage can be carried out *in situ*, and overall yields are, on average, greater than 80%.



The arylnitroacetylene can be selectively hydrogenated to the amino derivative, as shown above, using an unsupported ruthenium sulfide (RuS_2) catalyst. The yields and selectivities of the above reactions are greater than 90%. Other useful products made by this technology are *m*-bromophenylacetylene and *m*-diethynyl-benzene:

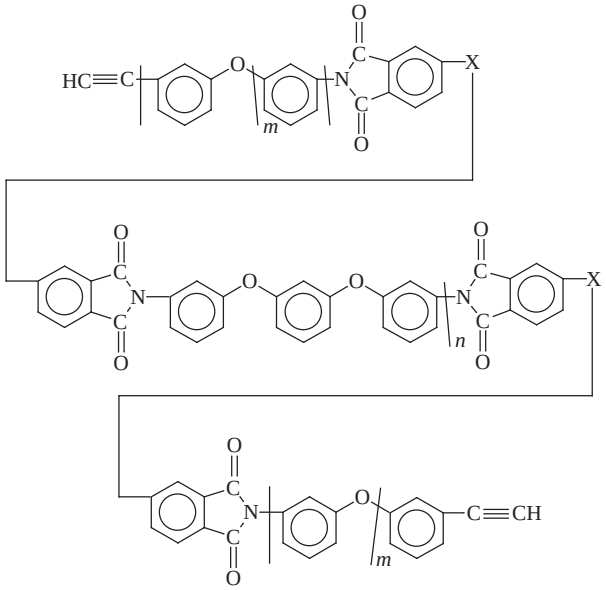


Base cleavage of the above compounds yields *m*-bromophenylacetylene and *m*-diethynylbenzene. These arylacetylene intermediates are used in the preparation of stable, high-temperature polymers, as discussed in Section XIII.C. The *m*-bromophenylacetylene can be used as a precursor to form *m*-aminophenylacetylene, which is an important acetylenic capping agent for polyimide polymers. This is an alternative method of preparing the *m*-amino derivative in place of catalytic hydrogenation, cited above.

C. Acetylene-Capped Polyimide Polymers

The work of N. Bilow and co-workers under sponsorship of the U.S. Air Force and Hughes Aircraft Company was responsible for novel, acetylene-terminated oligomers that could be cured in the absence of catalysts at 200°C to high-strength polymers. These acetylene-terminated polyimides cured readily without objectional evolution of volatiles such as water or alcohol to yield nonporous, void-free, high-strength polymers stable up to 370°C. Graphitereinforced composites cured with these acetylenic polyimides and subjected to aging in air at 320°C for 1000 hr had 75% of their flexural strength retained. The polymers were also observed to be superior laminating resins for 6 A1-4V titanium in forming titanium–titanium bonds stable at 500°F. The initial application of these resins was directed primarily at the aircraft and aerospace industries.

The basic polymer structure of some of these uncured acetylene-capped polyimides is as follows:



Polymer	*n*	X	*m*
HR-600	1	CO	0
HR-602	2	CO	0
HR-650	1	CO	1

The uncured polymers are made by reacting aromatic dianhydrides with aromatic diamines to form the polyamic

acid precursor, which is then reacted with *m*-aminophenylacetylene to end-cap the prepolymer. Azeotropic removal of water (refluxing benzene or toluene) converts the polyamic to polyimide acetylenic prepolymer. The latter is then cured at 200°C.

Gulf Chemical Company purchased the technology for these polymers and marketed the HR-600 product under the trade name Thermid 600. The applications recommended for the product included aircraft and missile structures, structural adhesives, low-friction bearings, nuclear-radiation-resistant parts, and circuit boards. The Thermid 600 line included a polyimide molding powder (MC-600), an *N*-methylpyrrolidone solution of the polyamic polymer (50% solids, LR-600), and a fast-drying alcohol solution of the polyamic ester form (Thermid AL-600). Gulf has sold this specialty acetylenic polymer business to National Starch.

D. Polyacetylene $(-CH-)_x$

McDiarmid and Shirakawa observed that the polymerization of acetylene in the vapor state on glass surfaces with Ziegler–Natta catalysts, followed by doping with a variety of additives such as halogens, antimony pentafluoride, perchloric acid, sodium, and lithium, resulted in greatly increased conductivity. A typical increase in conductivity from that of the original undoped polyacetylene to that of the doped polymer was of the order of 10^{-9} to 10^{3} (doped) Ω^{-1} cm^{-1}.

Polymerization of acetylene at room temperature yields a mixed cis–trans product, approximately 70–80% cis isomer. Polymerization at low temperature gives an all-cis polymer. Heating the cis isomer to 200°C for 1 hr converts it completely to the trans form. The conductivities of the undoped cis and trans isomers are essentially identical (cis, trans, 10^{-9}), while both doped isomers give conductivities up to 10^{3} Ω^{-1} cm^{-1} in resistivity.

Polyacetylene film is a fibrillar mat of randomly oriented strands. The dopants used are classified as electron acceptors (bromine, iodine, arsenic pentafluoride) or electron donors (lithium, sodium, potassium). A polymer treated with donor or acceptor dopants yields *n*-type or *p*-type semiconductors, respectively. Polyacetylene doped with SbF_5 or AsF_5 is a better conductor of electricity on a weight basis than mercury. The doped, random film mat can be stretched oriented by thermal and mechanical treatment to give conductivities as high as 2000 Ω^{-1} cm^{-1}. The conjugated polyene chain of polyacetylene and its formation of charge-transfer complexes with dopants are responsible for the greatly enhanced conductivities observed.

Polyacetylene has been given much publicity in publications, trade magazines, and the press. Applications for which it is claimed to have great potential are rechargable, lightweight automobile batteries and rechargable, portable energy-storage devices, solar batteries, radio and EMI shielding, fuel cells, Schottky devices, and wire-cable applications. Companies that have carried out research and commercial development activities with polyacetylene are Rohm and 2Haas, Allied, BASF, Xerox, IBM, GTE, and Showa Denko. Rohm and Haas has actually produced pilot quantities of undoped polyacetylene film for commercial sampling.

To date, no significant markets have developed for polyacetylene. The following problems are associated with its production and the utilization of doped polyacetylene:

1. Difficulty in melt processing and extruding
2. Deterioration and loss of conductivity in air and in the presence of moisture
3. Need for expensive processing equipment
4. Lack of compatibility with other polymeric materials

Polyacetylene, on a weight basis, is an outstanding conductive material, but its chemical environments and conditions of use must be carefully controlled. New dopants such as hexafluorophosphide ion (PF_6^-) render the polymer more stable to oxidation and hydrolysis, and further research along this line may be fruitful. Also, copolymerization with new acetylenic monomers or the formation of lower molecular weight, linear polyacetylenes may improve the melt processability of the polymer.

Polyacetylene doped with lithium (as the anode) is a highly conductive system, which because of polyacetylene's highly unsaturated (polyene) structure has the unique property of storing and releasing electrons. This property, together with the low unit weight of the cell, makes rechargable, portable batteries a commercial possibility. However, such a cell is highly vulnerable to air oxidation and traces of moisture. Also, the other difficulties cited above for doped polyacetylene have not yet been resolved. At some future time in new space–age applications using the inertness of outer space, polyacetylene may prove its worth.

The year 2000 Nobel Prize in Chemistry was awarded to Alan J. Heeger, Alan G. MacDiarmid, and Hideki Shirakawa for the discovery of and development of conductive polymers, which among the first were the polyactylenes.

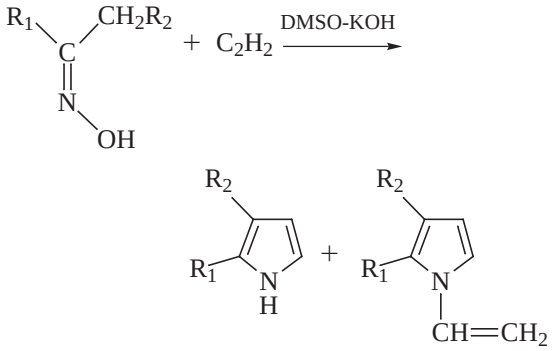
XIV. ACETYLENE RESEARCH IN RUSSIA

Russia has maintained a strong interest in acetylene chemistry since the discovery of the Favorsky reaction, in which

aldehydes and ketones in the presence of potassium hydroxide and polar solvents yield acetylenic alcohols and diols. A current site of Russian research is the Institute of Organic Chemistry of the Russia Academy of Sciences, at Irkutsk, Siberia. B. A. Trofimov and co-workers have utilized KOH–dimethyl sulfoxide (KOH–DMSO) as a superbase medium for a series of novel, acetylene-based reactions. Some of the reaction systems developed by Trofimov are described in this section.

A. Pyrroles and Vinylpyrroles from Acetylene and Ketoximes

The reaction of excess acetylene with ketoximes in KOH–DMSO medium under pressure at 70–140°C yields a mixture of pyrroles and *N*-vinylpyrroles. The reaction has been studied intensively with aliphatic, alicyclic, and aromatic ketoximes and has become known as the Trofimov reaction.



This method makes available many substituted pyrroles not readily synthesized. By using a limited or stoichiometric amount of acetylene, the pyrrole can be made the predominant product. The *N*-vinylpyrroles readily polymerize by a free-radical mechanism using azobisisobutyronitrile as the initiator.

B. Oligomers from Acetylene in KOH–NH_3–DMSO Media

Trofimov observed that low molecular weight oligomers of the polyene type could be formed from acetylene in a KOH–NH_3–DMSO system at 80–120°C under pressure (12–18 atm). The oligomers were deep brown to yellow solids, readily soluble in acetone, and insoluble in hexane and melted with decomposition at 300–400°C. The products by infrared examination contained C═O, C─O, and ─OH functionalities, but no nitrogen. Yields based on acetylene consumed averaged 75–88%. The oligomers were paramagnetic and conductive. Moderately polar amines such as triethylamine could be substituted for ammonia with equivalent results. However, the use of aniline required the addition of water to suppress nucleophilic addition of the amine across the triple bond.

Trofimov views oligomerization as proceeding simultaneously via the following routes:

$$\geqslant N: + HC{\equiv}CH \longrightarrow \geqslant N{-}\overset{+}{C}H{=}\bar{C}H \tag{13}$$

$$\geqslant N{-}\overset{+}{C}H{=}\bar{C}H + C_2H_2 \longrightarrow$$

$$|\geqslant\overset{+}{N}{-}CH{=}CH{-}(CH{=}CH)_{n-1}{-}CH{=}CH_2|OH^- \tag{14}$$

$$(CH_3)_2{-}SO + KOH \rightleftarrows$$

$$\left|CH_3{-}\underset{\underset{O}{\downarrow}}{S}{-}CH_2^- \longleftrightarrow CH_3{-}\underset{O\ -}{S}{-}CH_2\right| \overset{+}{K} + H_2O \tag{15}$$

I

$$\mathbf{I} + C_2H_2 \Longleftrightarrow K^+|(CH_3)_2{-}SO|C{=}\bar{C}H_2$$

II (16)

$$\mathbf{II} + nC_2H_2 \longrightarrow$$

$$HC{\equiv}C{-}|CH{=}CH|_{n-1}{-}CH{=}\bar{C}H{-}|(CH_3)_2SO|K^+ \tag{17}$$

$$OH^- + CH{\equiv}CH \longrightarrow HO{-}CH{=}\bar{C}H_2 \xrightarrow{C_2H_2}$$

$$\bar{C}H{=}CH{-}(CH{=}CH)_n{-}OH \tag{18}$$

The products are interesting because of their polyene, linear structure, solvent solubility, and conductive properties.

C. Reactions of Triads: S_8–KOH–DMSO, Se_8–KOH–DMSO, Te–KOH–HMPA

A number of interesting transformations reported by Trofimov and co-workers were based on the reaction of S_8, Se_8, and tellurium with acetylene in DMSO–KOH or HMPA (hexamethylphosphoramide)–KOH medium to yield the corresponding divinyl compounds:

$$C_2H_2 + \tfrac{1}{2}S_8 + KOH + H_2O \xrightarrow[80-120^\circ C]{DMSO} 2(H_2C{=}CH)_2{-}S$$

The reaction requires aprotic solvents such as DMSO and HMPA for satisfactory yields, with both solvents being approximately equal in performance (DMSO, 80%; HMPA, 76% yields). A small amount of water is required in the reaction to function as a proton transfer agent.

The reaction of the selenium triad (Se–KOH–DMSO) with acetylene yields a significant amount of 2-vinyloxy-1,3-butadiene, besides the expected divinyl selenide:

$$C_2H_2 + Se_8 + KOH + H_2O \longrightarrow (CH_2{=}CH)_2{-}Se + CH_2{=}\underset{\displaystyle O{-}CH{=}CH_2}{\underset{|}{C}}{-}CH{=}CH_2$$

26% 20%

This by-product results from the competitive hydration–trimerization of acetylene vs vinylation. Also, the lower yield of divinyl selenide is believed to be due to the rapid oxidation of selenide ions by DMSO.

The tellurium triad gives a yield (54%) intermediate between the sulfur- and selenium-based reactions:

$$C_2H_2 + Te + KOH + H_2O \xrightarrow[110-120^\circ C]{HMPA} (H_2C{=}CH)_2{-}Te$$

54%

The same reaction carried out in DMSO gives only a 30% yield. The most important commercial aspect of these reactions is the process for divinyl sulfide, since sulfur is an inexpensive abundant raw material.

Divinyl sulfide (DVS) is one of the most important products discussed, since it can be formed in 98% yield from hydrogen sulfide and acetylene:

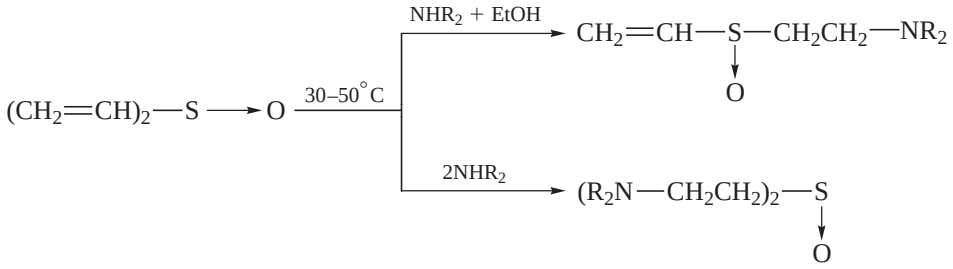
$$H_2S + C_2H_2 \xrightarrow[80-100^\circ C]{KOH-DMSO} \underset{\text{DVS (98\%)}}{(H_2C{=}CH)_2{-}S}$$

The reaction involving the addition of sulfide ion to acetylene may involve a one-step concerted mechanism, since all attempts to isolate monovinyl sulfide have failed. Either sodium sulfide or sulfur (cyclic S_8) can be used as the starting material, with the yields being somewhat lower (80%) than with H_2S. However, the use of elementary sulfur has an economic advantage.

DVS has been studied extensively as a starting material for other products via additions across its double bonds, cycloadditions, free-radical reactions, and electrophilic additions. A potentially important derivative is divinyl sulfoxide (DVSO) made by the oxidation of DVS with hydrogen peroxide:

$$DVS \xrightarrow[50^\circ C]{H_2O_2/H_2O} (CH_2{=}CH)_2{-}S \rightarrow O$$

While DVS is inert to nucleophiles, DVSO is quite active in Michael-type additions, leading to numerous mono- and di-adducts based on reactions with amines, alcohols, thiols, and active C–H compounds.



D. New Reactions and Chemicals Based on Sulfur and Acetylene

Trofimov's research in acetylene–sulfur chemistry includes the activation of anions and triple bonds with superbase systems, the reaction of sulfur nucleophiles with acetylene in superbase media, the reaction of substituted acetylenes with chalcogen nucleophiles (sulfur, selenium, tellurium) in superbasic media, and prospective applications. He predicts that in the future both acetylene and sulfur will be sources of inexpensive raw materials for making important acetylene–sulfur chemicals. Aprotic solvents such as DMSO and HMPA are required for the high yields obtained.

DVS readily polymerizes with free-radical catalysts and, as a cross-linking agent, yields macroreticular copolymers, which have excellent ion-exchange properties. DVSO has potential as a monomer in various copolymer systems and as a synthon for making new monovinyl derivatives.

E. Vinylox (2-Vinyloxyethoxymethyl Oxirane)

Vinylox is a monomer with interesting commercial potential. It is produced by partial vinylation of ethylene glycol followed by condensation of the resulting 2-vinyloxyethanol with epichlorohydrin.

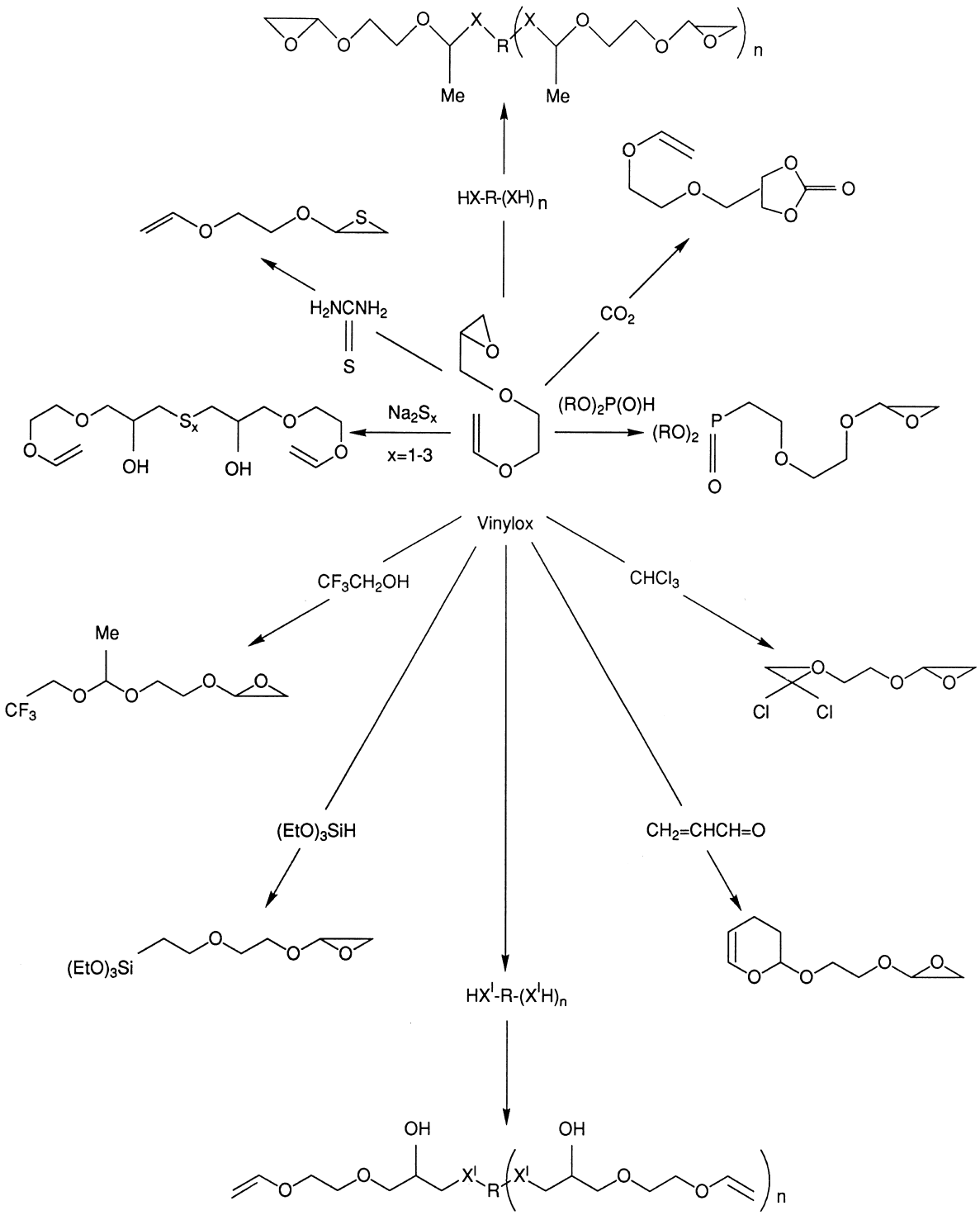


FIGURE 14 Vinylox-based derivatives. X = O, S, C(=O)—O; X^1 = O, S, N, C(=O)O; n = 0–7.

$$OH(CH_2)_2OH + C_2H_2 \xrightarrow{KOH} CH_2{=}CHO(CH_2)_2OH$$

$$CH_2{=}CHO(CH_2)_2OH + ClCH_2CH{-}CH_2 \quad (O)$$

$$\xrightarrow{NaOH} CH_2{=}CHO(CH_2)_2OCH_2CH{-}CH_2 \quad (O)$$

Vinylox

A process for 2-vinyloxyethoxymethyl oxirane has been developed and piloted in Russia by Trofimov *et al*. This proprietary process for Vinylox production is now being scaled up in Russia, since Vinylox has been shown by Russian scientists to be a highly active bifunctional monomer, reactive intermediate, and building block for diverse applications. It provides direct and reliable routes to numerous compounds with commercial promise such as epoxyacetals, vinyloxyethoxymethyl cyclocarbonates, and vinyloxyethoxymethyl thiranes. Some important

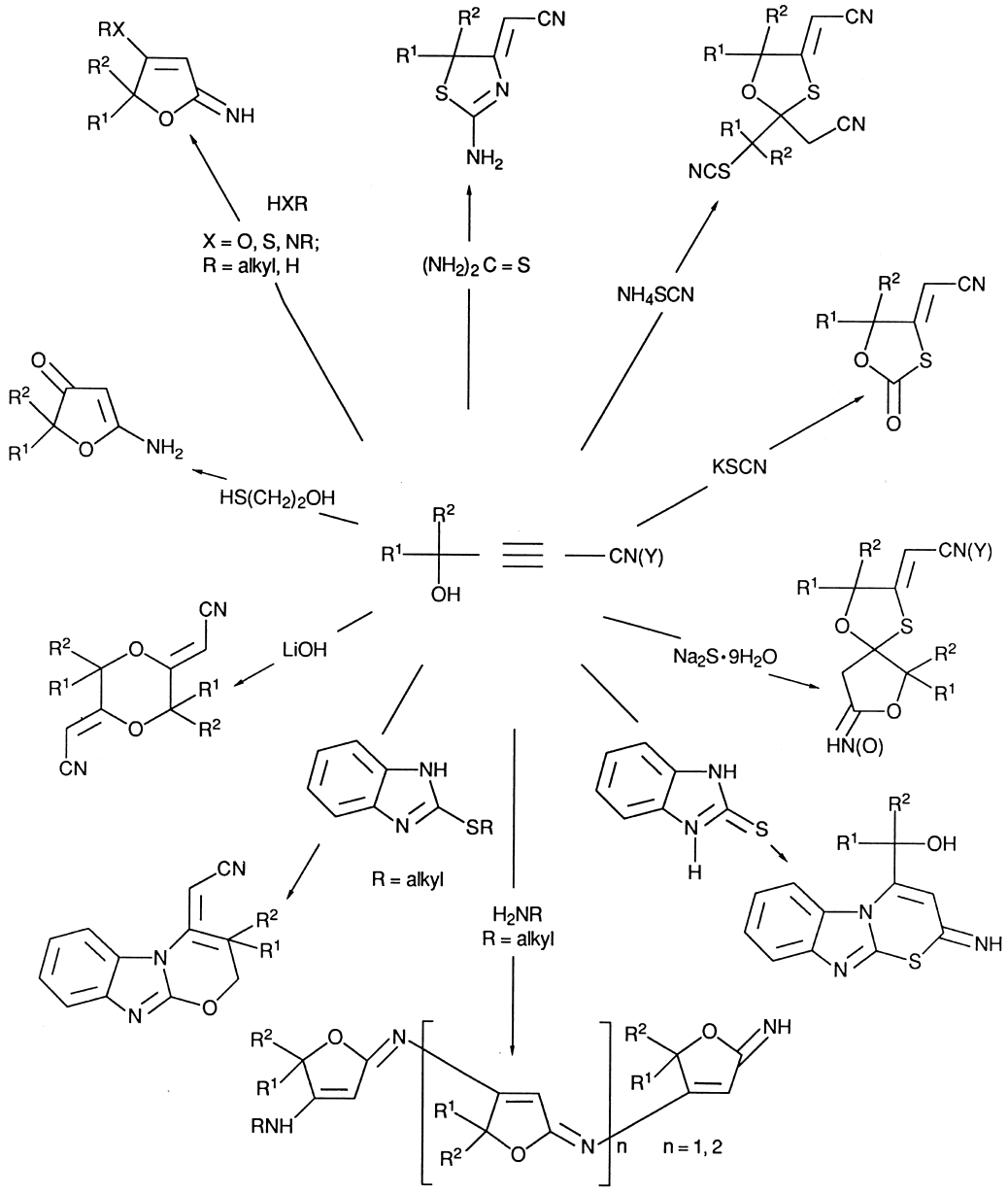


FIGURE 15 Hydroxyacetylenic esters and nitriles and their derivatives. Y = COOMe; $R^1 = R^2$ = alkyl, R^1–R^2–cycloaklylidene.

reactions and derivatives of Vinylox are shown in Fig. 14.

New epoxy resins of unique structures and high purity which are nontoxic, noncorrosive, and possess lower than expected viscosities are readily formed from this versatile monomer. The cured epoxy resin, in turn, exhibits higher strength and greater flexibility in a variety of composite applications. Polyols, glycols, carbohydrates, dicarboxylic and hydroxycarboxylic acids, polythiols, and other sulfur hydroxy compounds react readily with Vinylox to form the corresponding polyglycidyl derivatives. Vinylox and its derived epoxides are used as adhesives, active diluents, plasticizers, and modifiers for diverse epoxide materials. Small additions of Vinylox to synthetic rubbers enhance their strength, elasticity, and water-freeze resistance. This monomer and some of its derivatives exert a thermostabilizing effect on PVC formulations. Vinylox can be copolymerized with both radical initiators and anionic catalysts to give curable copolymers with *n*-butylvinyl ether, vinyl acetate, *N*-vinylpyrrolidone, and other typical monomers. With cationic catalysts, Vinylox forms both soluble and cross-linked polymers, depending on conditions.

Potentially important derivatives of Vinylox are 2-vinyloxyethoxy-methyl thirane (trade name Vinylox-S) and 3-(2-vinyloxyethoxy)-propylene-1,2-carbonate (trade name Cyclovin).

F. Hydroxyacetylenic Esters and Nitriles

Trofimov and co-workers have explored the chemistry of γ-hydroxy-α,β-acetylenic esters and acetylenic nitriles, prepared by the following routes.

1. $R_2{-}C(OH){-}C{\equiv}CH + CO + MeOH \xrightarrow[PdCl_2]{CuCl_2} R_2{-}C(OH){-}C{\equiv}C{-}CO_2Me$

2. $R_2{-}C(OH){-}C{\equiv}CH \xrightarrow{KOBr} R_2{-}C(OH){-}C{\equiv}C{-}Br \xrightarrow[DMSO]{CuCN} R_2{-}C(OH){-}C{\equiv}C{-}CN$

Esters and nitriles derived from tertiary acetylenic alcohols add smoothly to diverse O-, N-, and S-containing nucleophiles to yield hetrovinyl derivatives which undergo intramolecular cyclization to dihydrofurans and other heterocycles. From this chemistry, novel spirocyclicimines, lactones, and 1,3-oxathiolanes have been obtained from sulfur containing nucleophiles (sulfide and rhodanide ions), while acetylenic nitriles give new 1,3-thiazine systems and polyconjugated iminodihydrofurans. Figure 15 summarizes some of this novel chemistry, some derivatives of which exhibit pesticide and other bioactivity.

SEE ALSO THE FOLLOWING ARTICLES

CATALYSIS, INDUSTRIAL • PHARMACEUTICALS • PHYSICAL ORGANIC CHEMISTRY

BIBLIOGRAPHY

Bilow, N., Landis, A. L., and Boschan, R. H. (1978). U.S. Patents 4,098,767 and 4,100,138.
Bilow, N., Landis, A. L., and Boschan, R.H. (1982). *SAMPE J.*
Brandsma, L., Vasilevsky, S. F., and Verkruijsse, H. D. (1997). "Application of Transition Metal Catalysts in Organic Synthesis," Springer-Verlag, Berlin/New York.
Copenhaver, J. W., and Bigelow, M. H. (1949). "Acetylene and Carbon Monoxide Chemistry," pp. 246–294, Reinhold, New York.
Kirk, R. E., and Othmer, D. F. (1978). "Encyclopedia of Chemical Technology," 3rd ed., Vol. 1, pp. 192–276, Wiley-Interscience, New York.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1977). *J. Chem. Soc., Chem. Commun*. p. 578.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1978). *Appl. Phys. Lett*. **33,** 180.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1978). *J. Am. Chem. Soc*. **100,** 1013.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1979). *Coat. Plast. Prepr. Pap. Meet. Am. Chem. Soc., Div. Org. Coat. Plast. Chem*.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1979–1980). *Synth. Met*. **1,** 101–118.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1982). *Phys. Rev. B: Condens. Matter* [3] **26,** 2327–2330.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1984). *Mol. Cryst. Liq. Cryst*. **105,** 89–107.
McDiarmid, A. G., Park, Y. W., Heeger, A. J., and Shirakawa, H. (1984). *Polym. Prepr. Am. Chem. Soc., Div. Polym. Chem.*
Miller, S. A. (1965–1966). "Acetylene, Its Properties, Manufacture and Uses," Vols. 1, 2, Academic Press, New York.
Pasedach, H., Dimroth, P., and Schneider, K. (1961). German Patent 1,098,953.
Pasedach, H., Dimroth, P., and Schneider, K. (1962). German Patents 1,129,941 and 1,130,803.
Pasedach, H., Dimroth, P., and Schneider, K. (1963). German Patents 1,135,894 and 1,145,632.
Pasedach, H., Dimroth, P., and Schneider, K. (1963). U.S. Patent 3,082,216.
Roth, S. (1995). "One-Dimensional Metals," VCH, Weinbeim/New York.
Salaneck, W. R., Lundström, I., and Rånby, B., eds. (1993). "Nobel Symposium in Chemistry: Conjugated Polymers and Related Materials: The Interconnection of Chemical and Electrical Structure," Oxford Scientific, Oxford.
Selwitz, C. M., and Sabourn, E. T. (1978). U.S. Patents 4,128,588.
Selwitz, C. M., and Sabourn, E. T. (1980). U.S. Patents 4,204,078, 4,215,226, 4,216,341, 4,219,679, and 4,223,172.

Skvortsov, Yu. M., Fartysheva, O. M., Mal'kina, A. G., Sigalov, M. V., and Trofimov, B. A., *Zh. Org. Khim.* **22,** 255.
Skvortsov, Yu. M., Gritsa, A. I., Mal'kina, A. G., Sokolyanskaya, L. V., and Sigalov, M. V. *Sulfur Lett*. **6**(3), 87–92.
Skvortosov, Yu. M., Abramova, N. D., Andrijankova, L. V., Mal'kina, A. G., Kosinzina, E. I., Albanov, A. I., and Skvortsova, G. G. (1986). *Khim. Hetherocycl. Soed.* **10,** 1412–1415.
Skvortsov, Yu. M., Trofimov, B. A., Mal'kina, A. G., and Fartysheva, O. M. (1985). *Khim. Hetherocycl. Soed*. **12,** 1689–1690.
Tedeschi, R. J. (1963). *J. Org. Chem*. **28,** 1740.
Tedeschi, R. J. (1965). *Ind. Eng. Chem. Prod. Res. Dev*. **4,** 236.
Tedeschi, R. J. (1965). *J. Org. Chem*. **30,** 3045.
Tedeschi, R. J. (1968). *Ind. Eng. Chem. Process Des. Dev*. **7,** 303.
Tedeschi, R. J. (1969). *J. Org. Chem*. **34,** 435.
Tedeschi, R. J. (1969). U.S. Patents 3,441,621, 3,474,117, and 3,474,118.
Tedeschi, R. J. (1970). *Ind. Eng. Chem. Process Des. Dev*. **9,** 83.
Tedeschi, R. J. (1970). U.S. Patent 3,541,087.
Tedeschi, R. J. (1973). *Ann. N.Y. Acad. Sci*. **214,** 40–61.
Tedeschi, R. J. (1982). "Acetylene-Based Chemicals from Coal and Other Natural Resources," Dekker, New York.
Trofimov, B. A. (1980). "Heteroatomic Derivatives of Acetylene—New Polyfunctional Monomers, Reagents and Semi-Products," Nauka, Moscow.
Trofimov, B. A. (1980). *J. Polym. Sci. Polym. Chem. Ed.* **18,** 1547–1550.
Trofimov, B. A. (1982). "New Reactions and Chemicals Based on Sulfur and Acetylene," 10th Int. Symp. Org. Chem. Sulfur.
Trofimov, B. A. (1982). *Tetrahedron Lett*. **23**(48), 5063–5066.
Trofimov, B. A. (1982). *J. Appl. Chem. Russia* **51**(1), Pt. 1, 1091.
Trofimov, B. A. (1982). *Tetrahedron* **38**(5), 713–718.
Trofimov, B. A. (1984). *Sulfur Lett*. **2**(2), 55–58.
Trofimov, B. A., Stankevitch, V. K., and Belozerov, L. E. (1984). Russia Patent 1,129,208.
Trofimov, B. A., Nedolya, N. A., Khil'ko, M. Ya., Vyalykh, E. P., Mironov, G. S., and Moskvitchev, Yu. A. (1980, 1984), Russia Patents 751,811 and 1,074,881.
Trofimov, B. A., Nedolya, N. A., Vyalykh, E. P., and Figovskii, O. L. (1980). Rusia Patent 761,490.

Alkaloids

Armin Guggisberg
Manfred Hesse
The University of Zurich

GLOSSARY

Chemotaxonomy (chemosystematics) System of classifying plant species by their alkaloid (or other natural product) content.

Chromatography Method of separating the components of a mixture by distribution between a mobile phase (gas or liquid) and a stationary phase (e.g., silica gel, alumina).

Chromophore Functional group (nitro, carbon–carbon double bond, etc.) in an organic molecule that causes absorption of ultraviolet or visible light.

Configuration Spatial arrangement of atoms or groups of atoms around a central atom. (In structural formulas, C—H means the H is below, C—H means the H is above the paper plane).

Heterocyclic ring Ring in which one or more of the ring atoms are noncarbon atoms (e.g., oxygen, nitrogen).

Optical activity Capacity of some (chiral) compounds to rotate the plane of polarized light passing through. In the terminology of chiral compounds, the following symbols are used: (+), dextrorotatory, clockwise; (−), levorotatory, counterclockwise; ( ), indeterminate rotation or not to be found in literature.

ALKALOIDS are naturally occurring, nitrogen-containing organic compounds with the exception of amino acids, peptides, pterines and derivatives, purines and derivatives, amino sugars, and antibiotics. Clearly, the division of naturally occurring, nitrogen-containing substances into alkaloids and nonalkaloids is somewhat

arbitrary, and the boundary line is drawn in different positions by different authors.

I. ALKALOIDS IN HISTORY

Among the most notorious poisons in the history of civilization are the alkaloids. Very early in history, human beings learned to exploit these organic compounds for different purposes such as therapeutic treatment, defense, cosmetics, and getting food. They also used alkaloids to murder disagreeable fellow beings. Because they were not aware of the toxic effects of some special alkaloids, the use of these compounds led to small and great disasters.

During the course of civilization, knowledge of natural plant poisons and, accordingly, knowledge of poisonous plants was lost, especially in towns, which is documented by the increasing number of modern cases of poisoning.

The use of alkaloid-containing plants or plant parts with therapeutic or stimulating effects is mentioned in the oldest written documents. About 2700 BC, the Chinese Shen Lung described the drug *ma huang* and its use in medicine. The active component of this drug prepared from the plant *Ephedra chinensis* is the alkaloid ephedrine. In the "Papyrus Ebers" (1600 BC, Egypt), among the more than 80 medicinal plants or drugs described are the following alkaloid-containing drugs: hemlock (*Conium maculatum*), opium (from *Papaver somniferum*), *Ricinus*, and *Strychnos*. The European discovery of America made native American medicinal knowledge available in Europe and Asia. Drugs such as ipecacuanha (*Psychotria ipecacuanha*) and the Peruvian bark of *Cinchona* spp. stimulated medical interest because of their potent pharmacological activities. At about the same time, Europe became familiar with tabacco (*Nicotiana tabacum*). In the 17th and 18th centuries, symptoms of diseases such as fever (Peruvian bark), pain (opium), and constipation (*Ricinus*) were treated with plant products. Even diarrhea (ipecacuanha) and malaria (Peruvian bark) became curable. Tobacco and ephedra (against sleep) were used as stimulants.

Even in ancient times alkaloid-containing drugs such as opium, mandragora (*Mandragora officinalis*), and hen bane (*Hyoscyamus niger*) were used as anodynes in operations. Synthetic chemicals such as narcotics were not introduced until the 19th century.

Some of the drugs that were well known in the past are today being misused and abused and their use is therefore restricted. They are prepared mainly from alkaloid-containing plants, for example, opium and morphine from *Papaver somniferum*, cocaine from *Erythroxylum coca*, and mescaline from *Lophophora williamsii*. Repeated consumption of these drugs leads to mania and addiction (morphinism, cocainism, etc.).

In medieval Italy, extract of the deadly nightshade (*Atropa belladonna*) was once used to make women appear more beautiful (Italian: *bella donna*) by dilating the pupils of the eyes. Since prehistory, poisoned arrows and darts have been used to enhance the effect of a shot. The ancient Greek word *toxon* means "bow and arrow," and the Latin word *toxicum* for the poison used on arrows demonstrates the wide use of poisons in ancient times, for both war and hunting. In many regions of the world (East and West Africa, South America, southern Asia, Borneo, Java), extracts of alkaloid-containing plants were prepared to obtain the poison. In the northern part of South America, especially the upper basins of the Amazon and Orinoco rivers, the arrow poison (calabash and tubo curare) prepared from different species of the genus *Strychnos* caused respiratory arrest.

Another historical aspect of alkaloid use involves execution. For example, in 399 BC the Greek philosopher Socrates was sentenced to die by drinking a cup of hemlock (a liquid extract of the poisonous plant *Conium maculatum*). This was a common sentence at the time.

The biological effect of alkaloids also played a major role in the so-called Opium War (1840–1842). The British East India Company cultivated and monopolized opium preparation from *P. somniferum* in Bengal (India), and, since 1773, this company had exported an increasing amount of the drug to China. The importation of opium was forbidden in 1820 by the Chinese government. This act precipitated the Opium War in 1840. China was vanquished by England and, as a consequence, Hong Kong was annexed to England (Treaty of Nanjing). Furthermore, China was forced to open its ports to Western European nations and to the British opium trade.

Until the 19th century, an epidemic disease caused by bread poisoning was known in Western, Central, and Eastern Europe. It was called raphania or St. Anthony's fire. Its spread between the ninth and thirteenth centuries in France and Central Europe reached dramatic proportions since, especially for poor people, no substitute for the staple food, corn, existed. Today, we know that alkaloids of the fungus *Claviceps purpurea* were responsible for these pharmacological effects. The fungus grows mainly on rye and barley and sometimes on wheat and wild-growing grass, producing several alkaloids, the so-called ergot alkaloids, which contaminate the food by way of the flour. Consuming the poisoned food led to this disease, to multilations, and in many cases to death. Nevertheless, ergot has been used in medicine for many centuries (1582, herbal book of Lonizer) to enhance labor in women.

II. HISTORY OF ALKALOIDS

The year 1817 is regarded as the birth of alkaloid chemistry. In this year, the German pharmaceutical chemist

TABLE I Historical Data on Isolation, Structure Elucidation, and Synthesis of Some Well-Known Alkaloids

Alkaloid	Source	Isolation of the pure alkaloid	Elucidation of the correct structure	Determination of the absolute configuration	Synthesis
Morphine	Opium	1805	1925	1955	1952
Emetine	*Psychotria ipecacuanha*	1817	1948	1959	1950
Strychnine	*Strychnos nux-vomica*	1818	1946	1956	1954
Atropine	*Atropa belladonna*	1819	1901	1933	1903
Quinine	*Cinchona* bark(s)	1820	1907	1950	1944
Coniine	*Conium maculatum*	1827	1881	1932	1886

F. W. A. Sertürner isolated in pure state the compound morphine (or morphia), which he had already described in 1805. He identified morphine as the basic and active principle of the drug opium.

It was now possible to measure an exact dose of morphine owing to this first preparation of an active drug principle in the crystalline state. Before this time, the dosing of drugs by varying the composition was very dangerous. Sertürner's success stimulated others to isolate other alkaloids. Initially, the well-established medicinal plants or drugs were investigated, for example, *Psychotria ipecacuanha*, *S. nuxvomica*, *A. belladonna*, *Cinchona* bark, and *Conium maculatum*, to extract emetine, strychnine, atropine, quinine, and coniine, respectively (Table I). At that time pharmacists and chemists were concerned mainly with the purity, elemental composition, and subsequently the structures of these compounds. But in most cases, the structures of the alkaloids were too complicated to be elucidated because the necessary physical and chemical methods were in their infancy. As a result, structure elucidation in some cases required more than 100 years. Some of the methods for structure elucidation developed in the 19th century are still used (see Section VI).

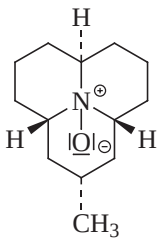
After the elucidation of its structure, an alkaloid is subsequently synthesized. Synthesis may offer proof of whether a structure is correct. The first synthesis of an alkaloid was that of coniine in 1886, 59 years after its first isolation from the plant.

The structure of a natural product is considered to be established as soon as its absolute configuration (i.e., the three-dimensional orientation of all atoms of the molecule) has been determined.

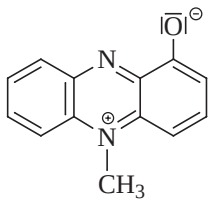
III. OCCURRENCE

The number of structurally different alkaloids has been estimated to be 6000. Most of them occur in flora, ~1% in animals, and not more than 0.5% in fungi and bacteria.

Coccinelline (from the European ladybird beetle *Coccinella septempunctata*) is an example of an animal alkaloid, whereas the deep-blue pyocyanine (from *Pseudomonas aeruginosa*) exemplifies bacterial alkaloids. It should be mentioned that, to date, not all living organisms have been investigated with respect to their alkaloid content. Therefore, the percentages mentioned are not definitive.



Coccinelline



Pyocyanine

Within the botanical classification system, it is interesting to consider the occurrence of alkaloid-containing plants. Botanists estimate the number of plant genera to be more than 20,000. (A genus is one order higher than a species.) Only 9% of all genera have alkaloid-containing species. These alkaloid-containing plants are not statistically distributed over the genera; rather, they occur most abundantly in genera belonging to the Dicotyledones and Monocotyledones of the Angiospermae (flowering plants). Table II shows the increasing number

TABLE II Alkaloid-Containing Plant Sections and Families Known in 1950 and 1985

Section	Class	Number of alkaloid-bearing plant families known	
		1950	1985
Angiospermae	Dicotyledones	28	120
	Monocotyledones	7	14
Gymnospermae		2	5
Pteridophytae		2	3

of plant families known to contain alkaloids and the dominance of the flowering plants. Alkaloids have rarely been found in the Gymnospermae and Pteridophytae.

It has been suggested that in 40% of all plant families, at least one alkaloid-containing species is known. On the other hand, there are plant families (e.g., Papaveraceae) in which all species contain alkaloids. Consequently, the distribution of the alkaloids in flora is very dissimilar.

Alkaloids may occur in several parts of a plant (e.g., roots, stem, bark, leaves, fruits, seeds). In some cases (e.g., *Papaver somniferum*, *A. belladonna*) alkaloids are isolated from all parts, whereas in other cases they are found in only one part (e.g., the alkaloid of *Aphelandra squarrosa* is found only in the roots). It should be noted that the alkaloid-containing part of the plant may not necessarily be the site of alkaloid formation.

The number of structurally different alkaloids varies from plant to plant. For example, *Catharanthus roseus* contains more than 100 alkaloids, whereas only 1 alkaloid has been detected in *A. squarrosa*. Quite often the ratio of the components of an alkaloid mixture is different in different parts of the plant. Other factors that influence the level and diversity of the alkaloid content are the age of the plant (e.g., investigated with *Adhatoda vasica*), the season (e.g., *P. somniferum*), the gathering time during the day (e.g., *Conium maculatum*), and the habitat (e.g., *Maytenus buxifolia*). The total amount of alkaloids in plants fluctuates between ~10% of quinine (*Cinchona* sp.) and $\sim 5 \times 10^{-6}$% of triabunnine in *Aristotelia peduncularis*, based on the dried weight of the drug.

IV. NOMENCLATURE

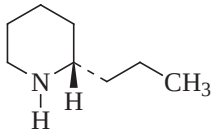
The system of nomenclature of alkaloids is similar to that of other natural products such as flavonoids and terpenoids. To date, no unique system has been developed. One of the major reasons is the vast number of different skeletal types. In most cases alkaloids are denoted by trivial names, which are derived from the (systematic) botanical name of the source plant. The compound name is derived either from the genus name (e.g., papaverine is isolated from different *Papaver* spp.) or the species name (harmaline is isolated from *Peganum harmala*). Some exceptions include pelletierine (referring to the name of the famous alkaloid chemist Pierre Joseph Pelletier), morphine (referring to the main physiological effect of an alkaloid; the Latin deity of dreams was Morpheus), and emetine (from Greek *emetikos*, "an emetic").

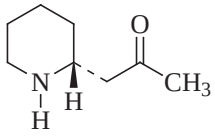
A common feature of nearly all alkaloid names is the ending "-ine" (English, French) or "-in" (German). In the case of several different alkaloids occurring in the same plant, the suffixes "-idine," "-anine," "-aline," "-inine," and so on are added to the common principal form.

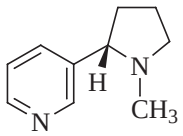
V. EXTRACTION, ISOLATION, PURIFICATION, AND ANALYSIS

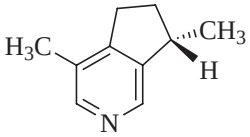
A. Physical and Chemical Features of Alkaloids

As with other organic molecules, the physical properties of alkaloids strongly depend on their molecular structures. Low molecular weight amines such as coniine (from *Conium maculatum*) and pelletierine (from *Punica granatum*) are colorless liquids. Other examples of liquid alkaloids are the *Nicotiana* alkaloids nicotine (from *N. tabacum*) and actinidine (from *Actinidia polygama*). Alkaloids of higher molecular weight—sometimes with additional oxygen functions—are usually colorless crystalline compounds, for example, papaverine (from *Papaver somniferum*). Several colored crystalline alkaloids are also known, such as the red-violet cryptolepine (from *Cryptolepis triangularis*).


(+)-Coniine

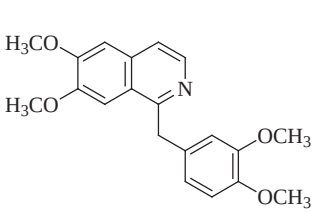

(+)-Pelletierine

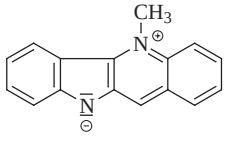

(−)-Nicotine


(+)-Actinidine

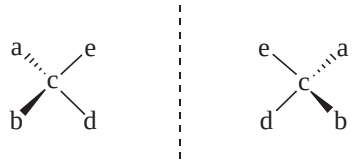
The majority of alkaloids are basic compounds. Whereas an alkaloid is usually soluble only in an organic solvent such as ether, ethanol, toluene, or chloroform and mostly insoluble in water, alkaloid salts (e.g., hydrochlorides, quaternary methochlorides) are soluble in water. Solubility in water is important for the therapeutic use of alkaloids.

Most isolated alkaloids are optically active. The plane of polarized light passing through a solution of an optically active compound is rotated either to the left (counterclockwise, levorotatory) or to the right (clockwise, dextrorotatory). The value of deflection is given, for


Papaverine
(m.p. 147–148°C)

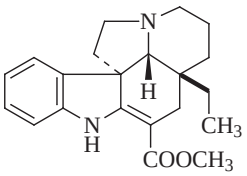

Cryptolepine
(m.p. 175–178°C)

example, as the $[\alpha]_D$ value. Optical activity is displayed by any molecule containing a carbon atom with four different substituents (asymmetric carbon atom).

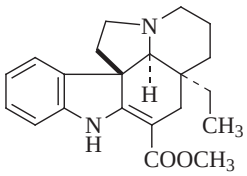


a, b, d, e = different substituents

Occasionally, the same alkaloid occurs as an equimolar mixture of levo- and dextrorotatory forms in the same plant. In this case the $[\alpha]_D$ value is zero. Such mixtures are called racemates. Vincadifformine was isolated from different plants in (−), (+), and (±) forms.



(+)-Vincadifformine

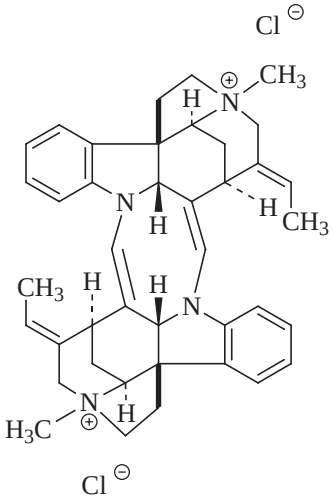


(−)-Vincadifformine

Source	$[\alpha]_D$
Amsonia tabernaemontana	+600°
Vinca difformis	±0°
V. minor	−540°
Hydrogenation of (−)-Tabersonine	−600°

B. Extraction and Isolation

Several methods have been established for extracting basic alkaloids from plant material. As an example a method frequently used for labscale extraction will be described. In plants, alkaloids usually occur as salts of common plant acids. The powdered drug is treated several times with a mixture of methanol and acetic acid, transforming the salts to alkaloid hydroacetates, which are soluble in methanol. The combined filtrates are evaporated, and the residue is taken up in 1% hydrochloric acid. Chlorophyll and other neutral or acidic compounds and neutral alkaloids can be extracted with an organic solvent such as ether or methylene chloride. The remaining acid solution is rendered basic with potassium carbonate, liberating the free alkaloids, which can be extracted with other or methylene chloride to yield the crude alkaloids. Alkaloids containing a quaternary nitrogen atom remain saltlike in the basic aqueous solution. They can be isolated by precipitation as water-insoluble salts such as picrates (with picric acid) or as Reineckates (with ammonium Reineckate) followed by displacement of the complex anion to chloride (e.g., through ion exchange). An example of a quaternary alkaloid is toxiferine dichloride from *Calebash curare*.



(−)-Toxiferine dichloride

C. Purification

Before preparative separation, it is necessary to examine the crude alkaloid extract by analytical thin-layer chromatography or by high-performance liquid chromatography to obtain preliminary information about the composition of the mixture. To detect alkaloids chromatographically, chromogenic reagents are used—for example, Schlittler's reagent (potassium iodoplatinate), Dragendorff's reagent (potassium bismuth iodide), or cerium(IV) sulfate/sulfuric acid.

If the crude alkaloid mixture consists mainly of one component, repeated crystallization from a suitable solvent may furnish the pure alkaloid. A more complex mixture must be separated by chromatography using silica gel, alumina, or cellulose powder as adsorbents and mixtures of organic solvents depending on the behavior of the individual alkaloids. Ion exchange and gel chromatography have also been successfully used. Other separation methods based on differences in partition between two immiscible liquids of functionally diverse alkaloids are sometimes preferred; among these are (Craig) countercurrent distribution and droplet countercurrent chromatography.

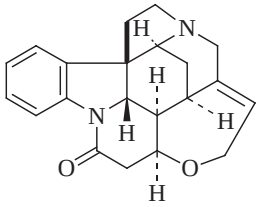
VI. STRUCTURE ELUCIDATION

The elucidation of the structure of an alkaloid requires the determination of its elemental composition and the bonded relationships of all constituent atoms in space. The starting point is a homogenous compound that cannot

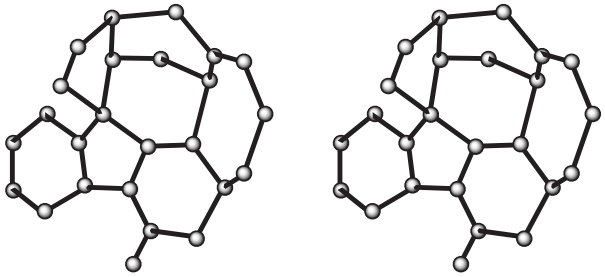
be separated into components by physical methods. The elemental composition of an alkaloid is determined by combustion analyses or by high-resolution mass spectrometry. Structure elucidation begins in earnest with the spectral investigation of the substance. Ultraviolet (UV) spectroscopy (light absorption in the UV region) yields information on the presence of chromophoric units such as aromatic and other unsaturated systems. Infrared spectroscopy identifies functional groups such as ketones, esters, amides, and hydroxyl groups and the degree of substitution of double bonds. Nuclear magnetic resonance (NMR) spectroscopy (especially of ^{1}H and ^{13}C nuclei) reveals information on the number and the electronic and stereochemical environment of the hydrogen and carbon atoms. Mass spectrometry provides the accurate molecular weight and the molecular formula through accurate mass measurement of the molecular ion and may provide structural information for the molecule by comparison of the fragmentation pattern with those of analogous systems.

In some cases it is possible to deduce the structure of an alkaloid using only one of these techniques, especially NMR and mass spectroscopy, but usually a combination of all methods is necessary. Chiroptical measurements (optical rotatory dispersion and circular dichroism) assist in determining the optical activity of the alkaloid. This reveals, by comparison with known compounds, the molecule in its three-dimensional orientation and may lead to the absolute configuration.

Finally, and independently of knowledge about elemental composition and spectral data, it is possible to determine the structure of a crystalline compound by X-ray diffraction analysis. An example of the results of such an analysis is the strychnine stereoprojection below. There is a possibility that the absolute configuration can be determined in this way.



(−)-Strychnine (absolute configuration)



Stereoprojection of a (−)-strychnine skeleton without hydrogen atoms (absolute configuration)

Only a small amount of a substance is required for spectroscopic analysis, particularly UV spectroscopy, mass spectroscopy, and X-ray diffraction (approx 10^{-5}, 10^{-9}, and 10^{-3} g, respectively).

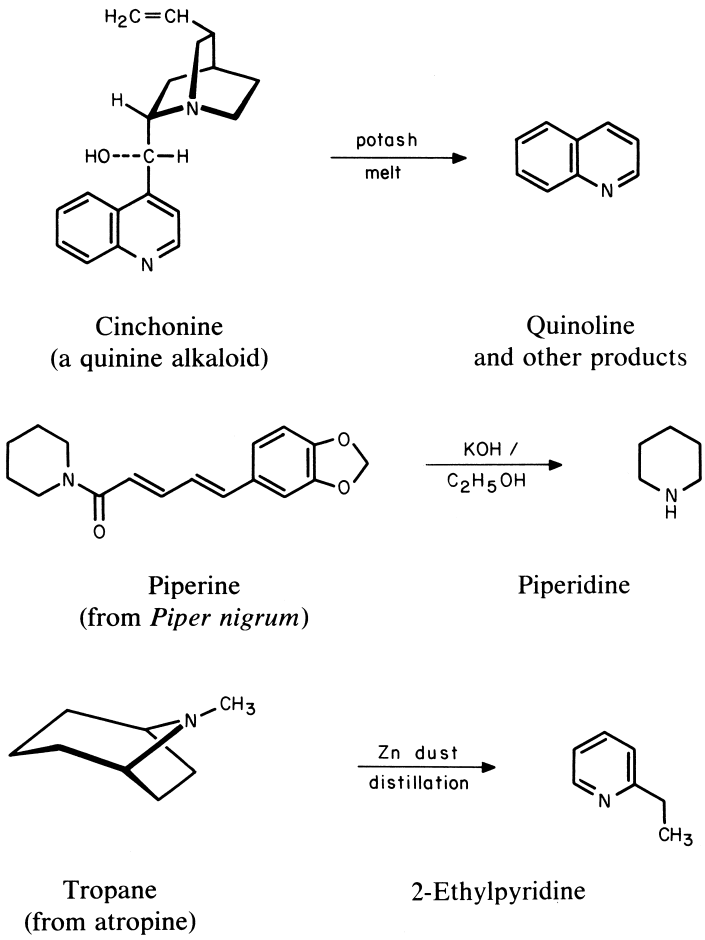
Chemical methods are also applied to the determination of structure. The major structural difference between alkaloids and most of the other natural products such as flavonoids and terpenoids is the presence of at least one nitrogen atom. Since 1820, chemically related structure elucidation of alkaloids has been based on specific reactions occurring due to the special reactivity of the nitrogen atom. The original reactions are of no value today; these include destructive distillation, potash melt, and zinc dust distillation. In many cases the destructive operations led to heterocyclic systems that were part of the alkaloids. Some of these heterocyclic systems were hitherto unknown, and the names given to them reflect the name of the parent alkaloid. Some examples illustrate this point.

The nomenclature of such compounds as quinoline, isoquinoline, and quinolizidine is derived from the first preparation of quinoline from quinine-type alkaloids (Scheme I)hhh. The name piperidine reflects its first preparation from piperine, a pepper alkaloid (*Piper nigrum*). The zinc dust distillation of the atropine-degradation product tropane furnished 2-ethylpyridine. Destructive reactions of this type are only of historical interest, since a number of alkaloid degradations led to heterocyclic systems that exhibited no structural relation to the alkaloid itself.

The most important reactions used in alkaloid chemistry are those that yield information about the neighborhood of the nitrogen atom(s). One of the most extensively used reactions is the Hofmann degradation of exhaustive methylation. Treatment of an alkaloid containing a basic nitrogen atom with methyl iodide leads to a quaternary ammonium salt. With base, the salt undergoes elimination of a β-proton and scission of the N—Cα bond to yield an olefin and a compound containing a tertiary nitrogen atom. Through repetition, until the nitrogen atom is extruded as $N(CH_3)_3$, this reaction permits the determination of the number of bonds between the nitrogen atom and the carbon skeleton of an alkaloid. In Table III the basic stages of the Hofmann degradation of three different alkaloid types are given. Other specific C—N degradation reactions are also occasionally used in structure elucidation of the alkaloids. Depending on the variation of alkaloid structures, oxidative, hydrolytic, reductive, and other reactions can also be applied.

VII. CLASSIFICATION

In the following survey of alkaloids, we consider, first, the nitrogen atom and its direct environment, and accordingly



SCHEME 1 Classical degradation reactions of alkaloids.

we classify the alkaloids on the basis of their nitrogen-containing structural features. In cases of alkaloids with more than one nitrogen atom, preference will be given to the most characteristic nitrogen function. Exceptions to this criterion will be allowed only if other structural elements are more typical, as in the case of steroidal, terpene, spermidine, spermine, and peptide alkaloids. Therefore, with respect to the location of the nitrogen atom, the alkaloids can be classified as follows:

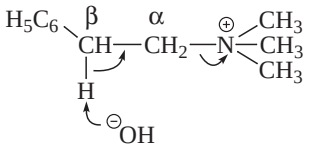
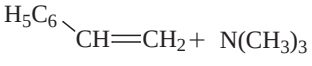
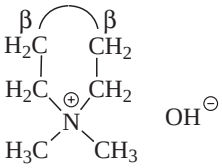
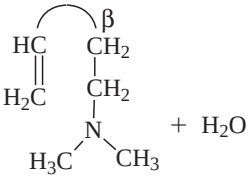
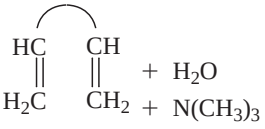
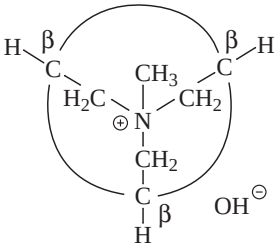
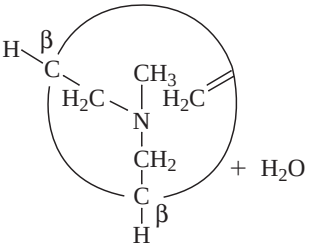
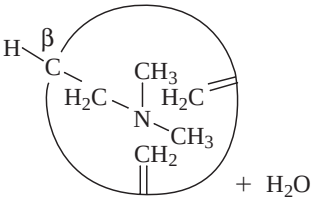
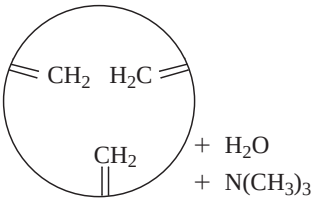
1. Heterocyclic alkaloids
2. Alkaloids with an exocyclic nitrogen atom and aliphatic amines
3. Putrescine, spermidine, and spermine alkaloids
4. Peptide alkaloids
5. Terpene and steroid alkaloids

According to this classification, the majority of alkaloids comprise the heterocyclic group, whereas the putrescine, spermidine, and spermine alkaloids form the smallest group.

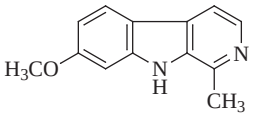
A. Heterocyclic Alkaloids

The representatives of this class have the nitrogen atom in a heterocyclic ring. Alkaloids containing one nitrogen atom as part of a five-membered ring are the so-called pyrrolidine alkaloids; an example is (+)-hygrine (from *Erythroxylum* sp.). The indole alkaloids form a large group within this class. The indole chromophore contains a five-membered ring and a benzene ring fused on one side. Some examples are vincadifformine, toxiferine dichloride (a bisindole alkaloid), strychnine, strictosidine, ajmalicine, and vincamine. Harmaline (*Peganum harmala*), (−)-physostigmine or (−)-eserine (*Physostigma venenosum*), and rutaecarpine (*Evodia rutaecarpa*) as well as cryptolepine and triabunnine are of different types. In pilocarpine (*Pilocarpus pennatifolius*), the

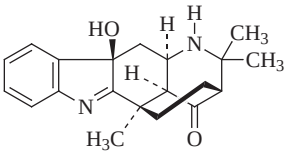
TABLE III Basic Stages of the Hofmann Degradation of Alkaloids

N-Methylation product of an alkaloid	First degradation product	Second degradation product	Third degradation product
	Repetition →	Repetition →	
	 Result: one bond between N atom and C skeleton		
		 Result: two bonds between N atom and C skeleton	
			 Result: three bonds between N atom and C skeleton

five-membered ring contains two nitrogen atoms, as in the amino acid histidine.



Harmaline

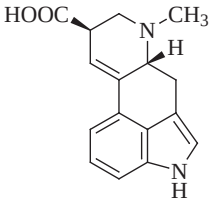


( )-Triabunnine

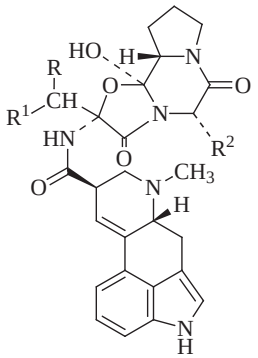
The peptide ergot alkaloids consist of one molecule of lysergic acid and a peptide unit. Typical examples are ergotamine and ergocornine, both isolated from *Claviceps purpurea*.

In a similar variation of alkaloids with a five-membered ring, the nitrogen atom is part of a six-membered ring. Examples are coniine, piperine, sedridine (*Sedum acre*), β-skytanthine (*Skytanthus acutus*), pelletierine, and actinidine. Nicotine consists of both a five- and a six-membered ring, each containing a nitrogen atom.

Through fusion of a benzene to a six-membered ring with one nitrogen atom, two ring systems can be formed:

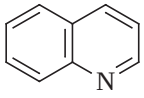


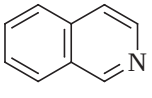
(+)-Lysergic acid



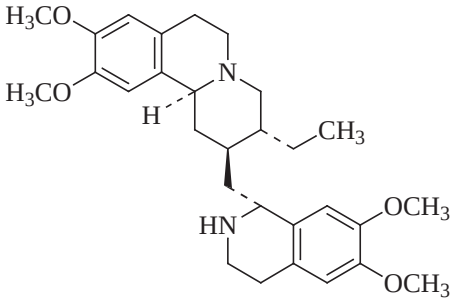
(−)-Ergotamine, R = R′ = H; $R^2 = CH_2—C_6H_5$
(−)-Ergocornine, R = R′ = CH_3; $R^2 = CH(CH_3)_2$

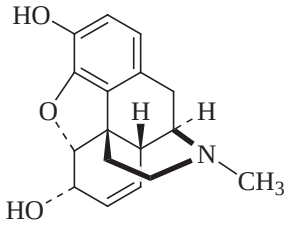
the quinoline and the isoquinoline system. An example of a quinoline alkaloid is the cinchona alkaloid cinchonine.


Quinoline

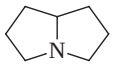

Isoquinoline

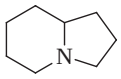
Natural isoquinoline alkaloids are very abundant; examples are papaverine, norlaudanosoline, emetine, and morphine. Although morphine cannot be classified as an isoquinoline alkaloid on the basis of its structure, biosynthetically it is derived from an isoquinoline alkaloid. Two nitrogen atoms in one six-membered ring are present in pyocyanine.

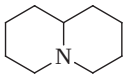

(−)-Emetine

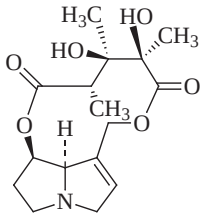

(−)-Morphine

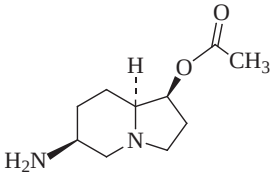
In quite a number of heterocyclic alkaloids, the nitrogen atom is common to two rings; some basic skeletal types and corresponding examples are the following:

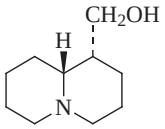

Pyrrolizidine


Indolizidine

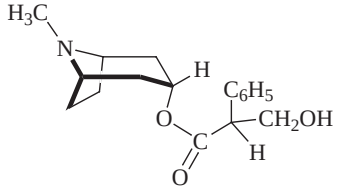

Quinolizidine


(−)-Monocrotaline
(from *Crotolaria* sp.)


( )-Slaframine
(from *Rhizoctonia leguminicola*)


(−)-Lupinine
(from *Anabasis* and *Lupinus* sp.)

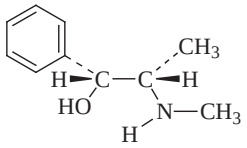

Tropane

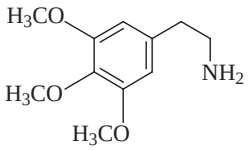

(+)-Atropine
(racemate of hyoscyamine)

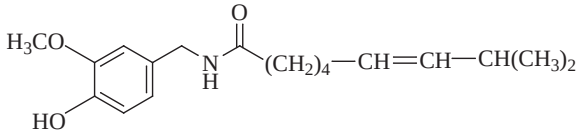
The nitrogen atom can also be common to three rings as in coccinelline. In addition to the heterocyclic systems mentioned above, some others occur as parts of alkaloids.

B. Alkaloids with an Exocyclic Nitrogen Atom and Aliphatic Amines

Ephedrine and mescaline are two well-known representatives of this class. Capsaicine was isolated from *Piper* and *Capsicum* species.

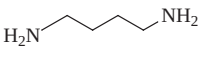

(−)-Ephedrine

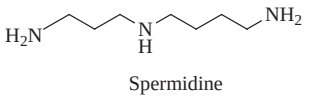

Mescaline

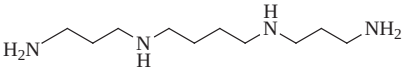

Capsaicine

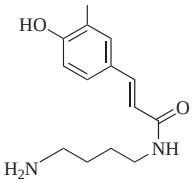
C. Putrescine, Spermidine, and Spermine Alkaloids

The alkaloids in this class contain putrescine, spermidine, and spermine as the basic backbone; typical examples are paucine, inandenin-12-one, and aphelandrine, respectively.

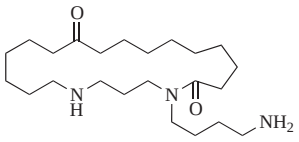

Putrescine


Spermidine

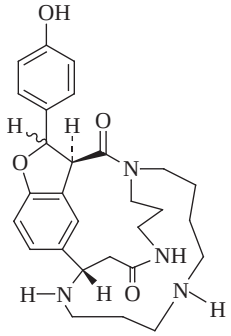

Spermine



Paucine
(from *Pentaclethra sp.*)



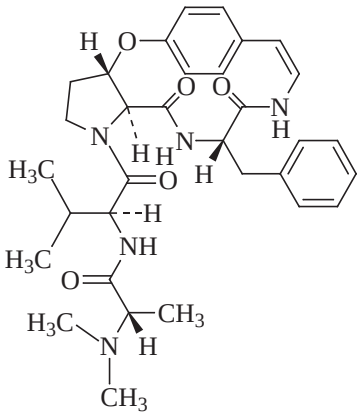
Inandenin-12-one
(from *Oncinotis inandensis*)



(+)-Aphelandrine
(from *Aphelandra squarrosa*)

D. Peptide Alkaloids

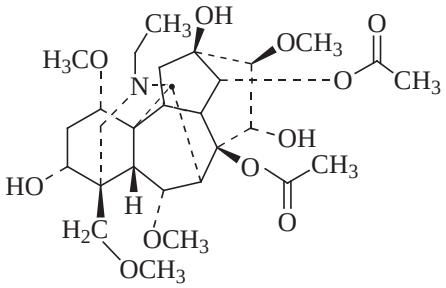
Besides a (cyclic) peptide, the peptide alkaloids contain an additional basic nitrogen atom, as demonstrated by mauritine A built up from the L-amino acids valine, *trans*-3-hydroxyproline, *N*,*N*-dimethylalanine, and phenylalanine, together with *p*-hydroxystyrylamine as the amino component. This group of alkaloids are also called rhamnaceous alkaloids owing to their predominant occurrence in the plant family of Rhamnaceae.



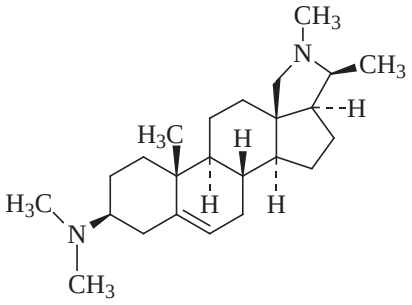
(−)-Mauritine A
(from *Zizyphus mauritiana*)

E. Terpene and Steroidal Alkaloids

Terpenes and steroids can be part of an alkaloid skeleton that contains in addition at least one nitrogen atom. Besides monoterpene (e.g., β-skytanthine), sesquiterpene, diterpene (e.g., aconitine, the poison of monk's hood), triterpene alkaloids, and steroidal alkaloids (e.g., conessine) are known.



(+)-Aconitine
(from *Aconitum napellum*)

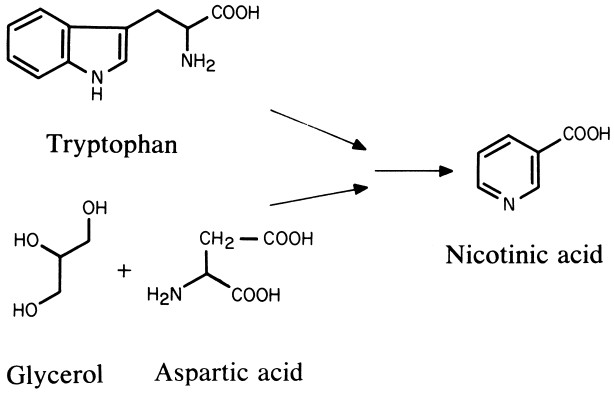


(−)-Conessine
(From *Holarrhena* species)

Other classification systems for alkaloids have been proposed. In the biogenetic classification, the class designations are derived from the amino acids or other fundamental units that are considered precursors of alkaloid formation in plants. This system has some advantages. As demonstrated in this article, even the chemical classification systems usually require biogenetic arguments in the case of special alkaloids (e.g., morphine). However, it is known from various organisms that they can biosynthesize the same alkaloid from different amino acids. In Scheme 2, the biogenesis of nicotinic acid is presented. The precursor in the first pathway is tryptophan; that in the second is aspartic acid. Although many biosynthetic pathways have been established experimentally, some biogenetic proposals are only speculative. Considering these disadvantages, it seems more plausible to use the classification system based on the alkaloid structures.

VIII. BIOSYNTHESIS

The alkaloid chemist is interested in the identity of the natural building blocks of alkaloids. On the basis of many experiments it has been demonstrated that for the

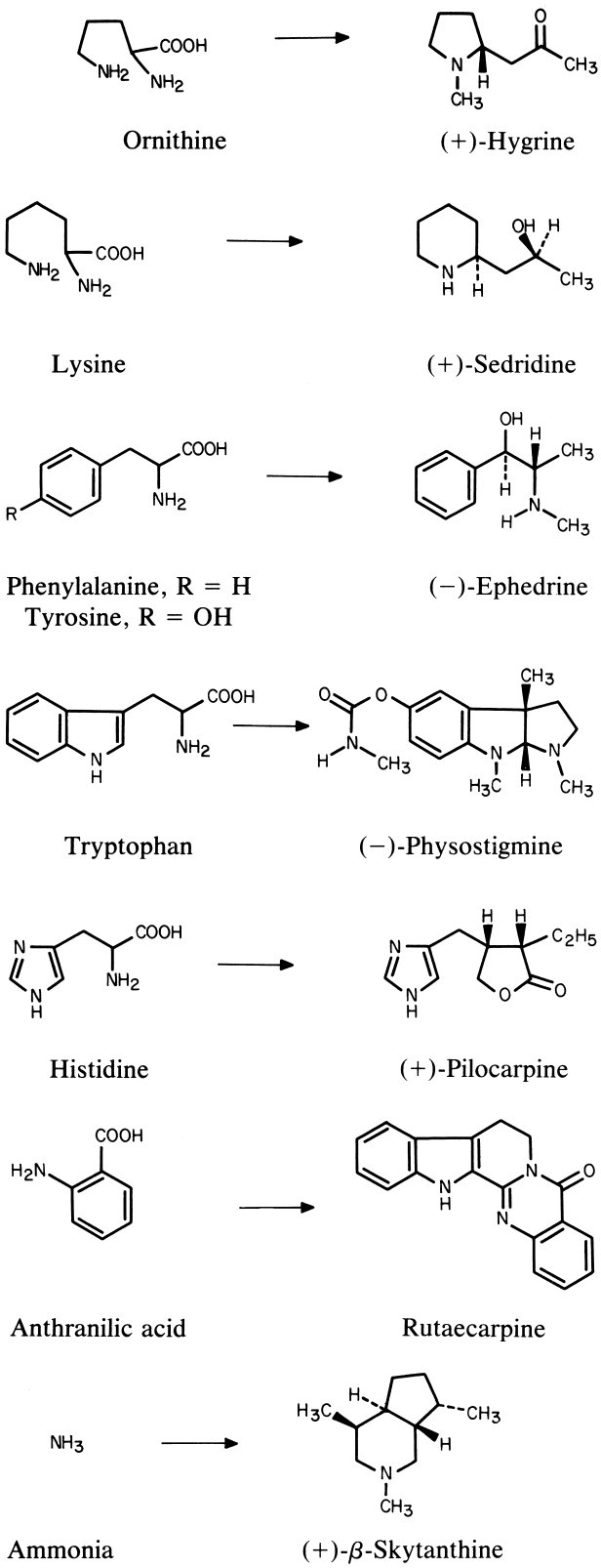


SCHEME 2 Two different pathways in the biosynthesis of nicotinic acid.

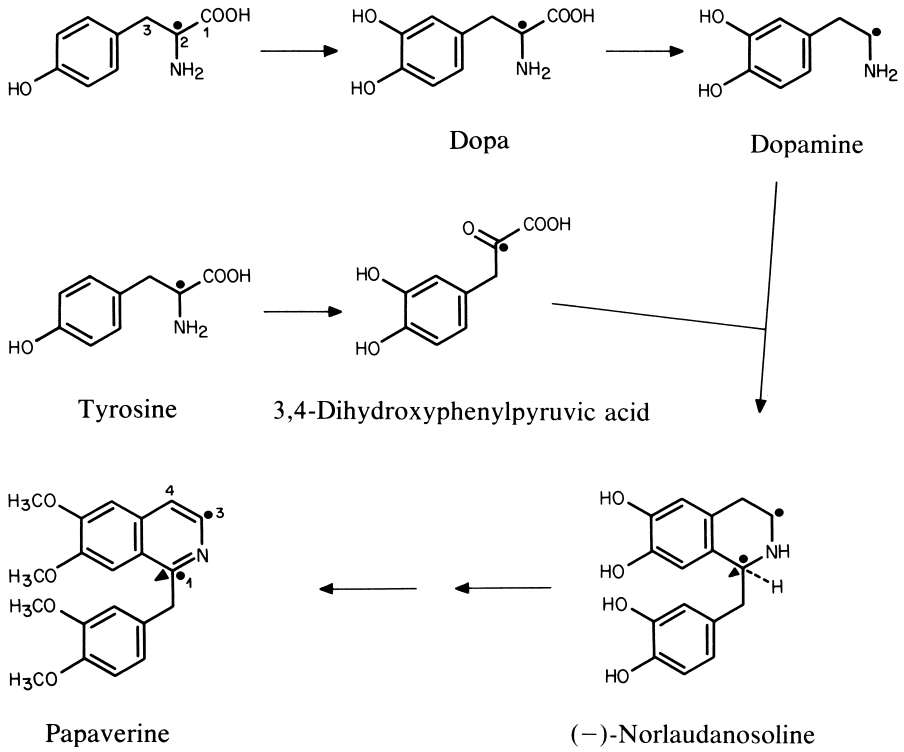
large variety of alkaloids only a few building blocks are needed. Structural variety is generated by special enzymatic systems in plant and animal cells. As expected, amino acids play the most important role among the nitrogen-containing precursors. But only a few amino acids act as precursors in alkaloid biosynthesis. Scheme 3 lists the important amino acids (without specification of their absolute configuration) together with their corresponding alkaloid derivatives. In the case of rutaecarpine, tryptophan is required in addition to anthranilic acid as a second amino acid precursor.

The same amino acids that are the precursors of the alkaloid examples given in Scheme 3 act as precursors for most of the other alkaloids. The remaining plant bases are derived by the addition of ammonia or an alkylamine to nonbasic precursors. Many of the nitrogen-building blocks of alkaloids are terpenoids, fatty acids, and cinnamic acid derivatives. Small carbon fragments, such as methyl acetate, and carbonate complete the peripheral functionalities. Carbon–carbon bond formation, oxidation, and reduction reactions are typical enzyme reactions.

It is possible to establish a biogenetic pathway in a living organism by using isotopic tracers. In feeding experiments, specifically labeled precursors are introduced into a plant or microbial broth. The labeling may be with either radioactive (e.g., ^{3}H, ^{14}C) or stable (e.g., ^{13}C) isotopes. Some difficulties are associated with this procedure. The labeled precursor must be synthesized with the labels in the known positions, and the amount of the label at each position established. Once labeled precursor has been incorporated, the product must be isolated in a pure state, the amount of label present in the whole molecule determined, and the precise location of the label established by chemical degradation of the molecule to pure products, the radioactivity of which is then measured. In the case of stable-isotope labeling, ^{13}C-NMR spectroscopy is



SCHEME 3 Examples of alkaloids and their corrsponding amino acid precursors.



SCHEME 4 Main pathways in the biosynthesis of papaverine from tyrosine in *Papaver somniferum* L (● *and* ▲ = ^{14}C).

necessary for the detection of the marked positions. From such experiments it can be ascertained whether a proposed compound is a precursor of an alkaloid. It is important to note that the results are valid only for the specific plant product of a specific plant because different pathways for the same product are known.

As an example, the biogenesis of the alkaloid papaverine from *Papaver somniferum* is given in Scheme 4. Tyrosine is the precursor of papaverine in *P. somniferum* since introducing the labeled tyrosine in position 2 by ^{14}C (=[2-^{14}C]tyrosine) into the plant resulted in incorporation into papaverine at positions C-1 and C-3. Therefore, two molecules of tyrosine are transformed into one molecule of papaverine. Several chemical reactions are involved in the transformation of tyrosine to papaverine: for example, in tyrosine the aromatic nucleus is substituted only by one oxygen atom, tyrosine contains a carbocyclic acid residue that is not present in the product, and in papaverine only one nitrogen atom is present, compared with the two in the precursor. Detailed analysis of the pathway, again using labeled precursors, showed that tyrosine is first hydroxylated in plant cells into 3,4-dihydroxyphenylalanine (Dopa). Dopa is decarboxylated (loss of CO_2) to give dopamine. Feeding experiments with dopamine have clearly demonstrated that only one molecule of dopamine is incorporated into papaverine. The conclusion of this experiment is that tyrosine plays a double role. It is hydroxylated and decarboxylated to form dopamine, and it is transformed into the intermediate 3,4-dihydroxyphenylpyruvic acid. The condensation reaction between an amine and a ketone, as well as the cyclization of the intermediate imine, is a well-known organic synthetic process. In case of the biosynthesis of papaverine, norlaudanosoline is the product of condensation, cyclization, and decarboxylation. With the aid of labeling experiments, it has been verified that norlaudanosoline is indeed the precursor of papaverine. Furthermore, it has been shown that the *O*-methylation of (−)-norlaudanosoline proceeds stepwise and partly parallel with dehydrogenation to papaverine.

IX. CHEMOTAXONOMY

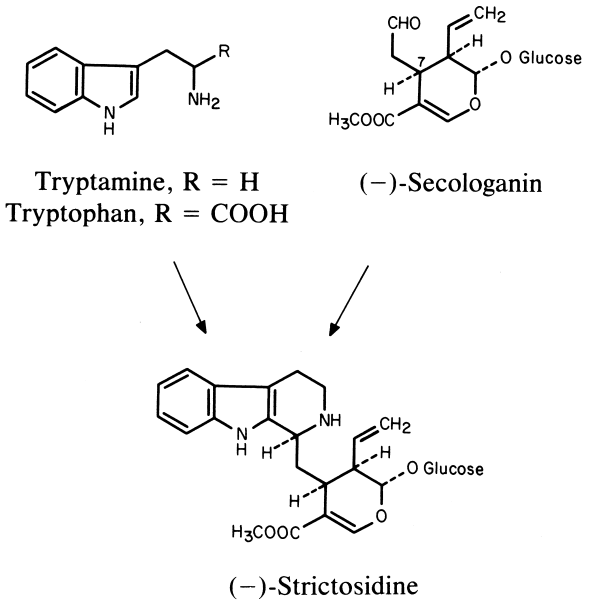
As mentioned above, the occurrence of alkaloids in plants is strongly related to the plant families. For a long time, it has been known that closely related plants contain the same structurally similar alkaloids. For example, the two tropan alkaloids atropine and hyoscyamine were isolated from the nightshade plants (Solanaceae) *Atropa beiladonna* (deadly nightshade), *Datura stramonium* (thorn apple), and *Mandragora officinarium*

(mandrake). Therefore, a classification of plant species based on the structures of their alkaloids and other chemical ingredients has been attempted. Such a system of classification of plant species by chemical compounds is known as plant chemosystematics or chemotaxonomy.

To investigate a group of alkaloids in the chemosystematic sense it is necessary to have a large body of statistical material available; otherwise, the results will be of uncertain value.

The distribution of indole alkaloids in plant species has been quite well investigated. The number of structurally known alkaloids is ~1200. They have been isolated from more than 400 different plant species. Some alkaloids were isolated from only one species; others were from many different species. A total of 1200 indole alkaloids have been isolated nearly 3500 times. These data underline the importance of this class of alkaloids for chemosystematic studies. To establish a botanical classification system on the structural basis of the indole alkaloids plants contain, it is important to investigate their biogenetic relationships.

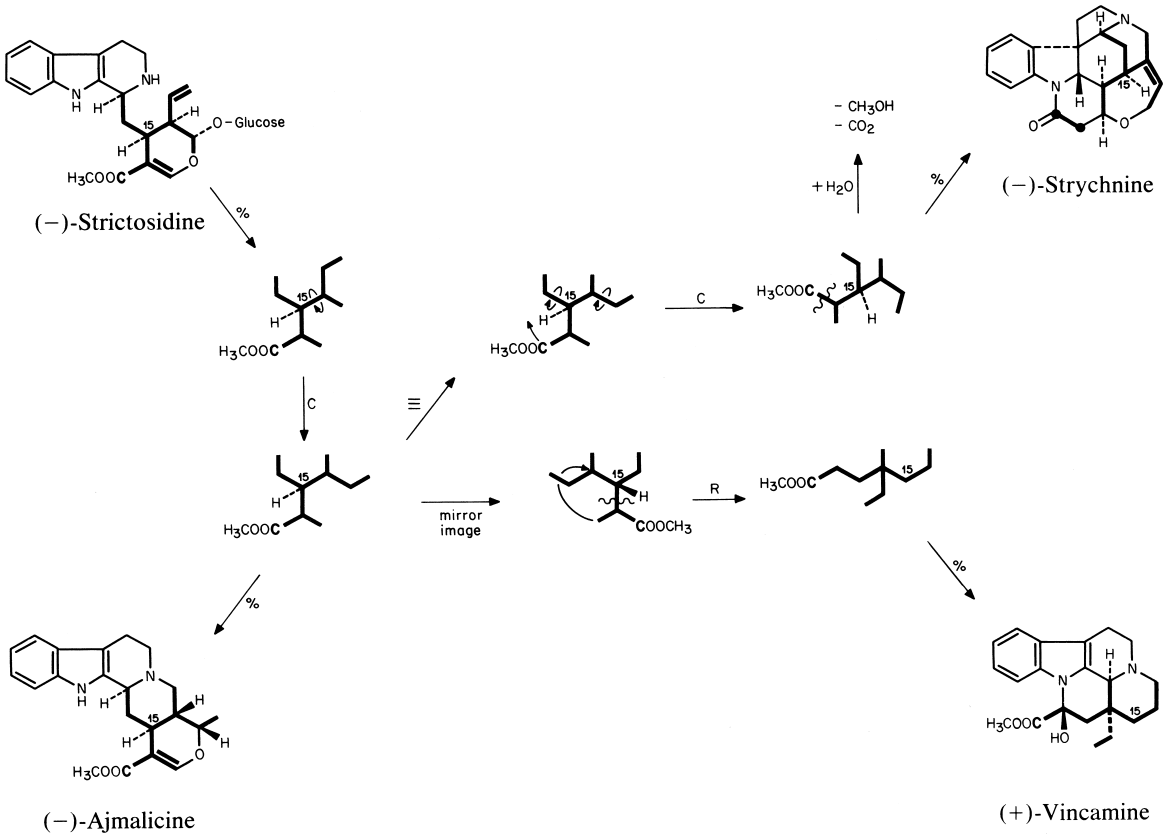
The above-mentioned 1200 indole alkaloids taken into consideration are built up from tryptamine or tryptophan (the basic components) and the monoterpene derivative secologanin. It was shown that tryptamine and secologanin are directly transformed to the alkaloid strictosidine (Scheme 5). Most of the 1200 indole alkaloids are derivatives of strictosidine. They are distinguished from strictosidine by variations in the functional groups (e.g., —H → —OH; NH → N—CH_3) and/or loss of carbon atoms (e.g., —$COOCH_3$ → —H) and/or opening or forming of chemical bonds. In Scheme 6 examples of structural changes of strictosidine leading to the indole alkaloids ajmalicine, strychnine, and vincamine are given. The structural changes occur only in the secologanin part of strictosidine: in all cases the tryptamine part remains unrearranged. Therefore, to emphasize the changes in the secologanin part, the carbon skeleton of the latter is presented in heavy lines in the scheme. These changes can occur by rotations around or rearrangements of various carbon–carbon bonds. Changes of oxygen functionality and double bonds and so on are not considered. The carbon skeleton of ajmalicine can be derived from that of strictosidine by rotation of one carbon–carbon bond. Two additional rotations are necessary to explain the formation of the carbon skeleton of strychnine. In addition, the $COOCH_3$ group should be eliminated and two additional carbon atoms (as an active acetate unit) have to be introduced into the molecule to reach strychnine. To explain the formation of the carbon skeleton of vincamine from that of ajmalicine, a rearrangement of a three-carbon chain is necessary. Strictosidine, ajmalicine, and strychnine are examples of indole alkaloids with a so-called nonrearranged carbon skeleton. All of the alkaloids with a nonrearranged carbon skeleton have the same absolute configuration at C-15 (marked in Scheme 6) as in secologanin at C-7 (Scheme 5). In the rearranged types, the chirality of C-15 disappears.



SCHEME 5 The indole alkaloid strictosidine is built up from the base trytamine and the monoterpene secologanin.

Nearly all (more than 99%) of the 1200 indole alkaloids mentioned above were isolated from plant species belonging to only three plant families: Apocynaceae, Loganiaceae, and Rubiaceae. It is remarkable that alkaloids of the strictosidine and ajmalicine types were isolated from the plants of all three families. Alkaloids of the strychnine group are known to occur only in the Loganiaceae. Alkaloids with a rearranged secologanin skeleton (e.g., vincamine) occur only in the Apocynaceae. Characteristic alkaloids of the Rubiaceae are ajmalicine-type compounds in various oxidation states.

The chemotaxonomic conclusion from these results are as follows. The enzyme system generating strictosidine-type alkaloids from tryptamine and secologanin must be the same in all three plant families (strictosidine synthetase). The most primitive plant family seems to be the Loganiaceae because the species of this family produce only the unrearranged skeleton types. Their special ability consists in the incorporation of an additional C_2 unit. To rearrange a carbon skeleton, more advanced enzyme systems are necessary. Therefore, it is deduced that the Apocynaceae are more developed than the Loganiaceae. For other reasons not mentioned, Apocynaceae and Rubiaceae are thought to be equivalent. Besides special oxidation



SCHEME 6 Structural relation of strictosidine to other types of indole alkaloids. Carbon skeleton of the secologanin unit is given in heavy lines. % denotes equivalent to: ∿, cleavage; R, rearrangement of carbon–carbon bonds; C, conformation changes by rotation around carbon–carbon bonds.

reactions, they can use tyramine (decarboxylated tyrosine; Scheme 3) instead of tryptamine in the alkaloid formation (e.g., emetine). A more detailed analysis of alkaloid structures in relation to plant families, subfamilies, tribes, genera, and species led to a proper classification of some plant species that had been incorrectly placed in the botanical classification system.

X. ROLE

A. In Natural Sources

Alkaloids can be described as products of secondary plant metabolism like other complex natural compounds such as flavonoids, terpenoids, and steroids. The role of alkaloids in alkaloid-containing plants is still a matter of speculation. For a long time, it has generally been accepted that alkaloids are end products of metabolism and are waste products of the plant. However, more recent investigations have shown that alkaloids are involved in a dynamic process; for example, the amount of the alkaloid coniine in *Conium maculatum* varies during a single day as well as in the course of the development of the plant. There have been similar observations of the atropine alkaloids in *A. belladonna* and the opium alkaloids in *P. somniferum*. In isotopic labeling experiments it was demonstrated that the turnover rate (half-life) of alkaloids in plants is very fast. This clearly shows that plants do metabolize alkaloids, and therefore the theory that alkaloids are metabolic end products is no longer tenable. However, whether alkaloids are of any use to the plant is still debatable since in only a few cases has it been shown that alkaloid-containing plants are protected against consumption by animals. On the other hand, both alkaloid-free and alkaloid-containing plants can be attacked by parasites, fungi, and bacteria.

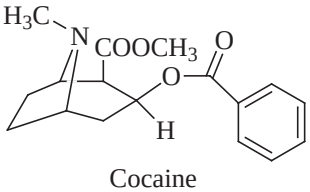
In contrast to their role in plants, the function of alkaloids in insects is much more established; for example, coccinelline, the tricyclic alkaloid *N*-oxide, is produced by beetles of the Coccinellidae and is used as a defensive allomone against predators.

TABLE IV Pharmacological Properties of Some Alkaloids

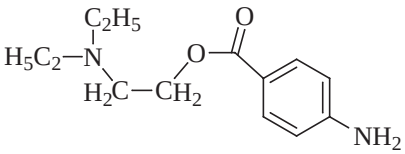
Alkaloid	Pharmacological action
Aconitine	Antipyretic, antineuralgic
Ajmalicine	Antihistaminic
Ajmaline	Antiadrenergic, antiarrhythmic
Atropine	Anticholinergic, mydriatic, antispasmodic
Camptothecine	Antitumor activity
Cocaine	Local anesthetic, vasoconstrictor
Codeine	Antitussive
Emetine	Emetic, amebicide
Ephedrine	Sympathomimetic
Ergot alkaloids	α-Sympatholytic, vasodilator, antihypertensive
Homoharringtonine	Antitumor activity
Morphine	Analgesic, narcotic
Papaverine	Antispasmodic
Physostigmine	Parasympathomimetic, miotic
Pilocarpine	Parasympathomimetic, diaphoretic, miotic
Quinidine	Antiarrhythmic
Quinine	Antimalarial
Reserpine	Antihypertensive
Scopolamine	Mydriatic, parasympatholytic
Tubocurarine chloride	Muscle relaxant
Vincamine	Vasodilator, antihypertensive
Vincaleucoblastine	Antineoplastic
Vincristine	Antineoplastic

B. Alkaloid Structures in Drug Design

The pharmacological importance of alkaloids is enormous. Since the discovery of morphine by Sertürner, many plant alkaloids have been used directly or by molecular modification for therapeutic purposes. Even today a number of alkaloids (or their hydro salts) are used clinically. Examples are given in Table IV; their structures are shown in Scheme 8.



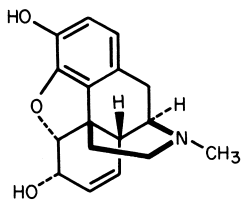
Cocaine



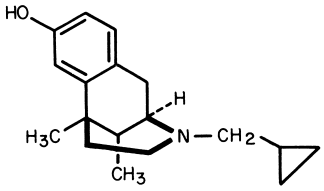
Procaine

Other useful therapeutic agents are derivatives of alkaloids with minor changes in the molecular structure, such as 2,3-dehydroemetine (amebicide), strychnine *N*-oxide (analeptic), and *N*,*N*′-diallylnortoxiferine dichloride [potent muscle relaxant; it differs from toxiferine dichloride in the substituents of two nitrogen atoms: (N)—CH_2—CH═CH_2 instead of (N)—CH_3]. Finally, another group of compounds used therapeutically differ substantially in structure from the natural alkaloids. In their analgesic action cocaine and procaine are similar. The analgesic activity of morphine greatly stimulated molecular modification of its structure, even before all the details were known. Cyclazocine is considered to be ~40 times more potent analgetically than morphine given subcutaneously or orally. Because of side effects it is unacceptable as a general analgesic. Etorphine, another modified morphine, is 100–80,000 times more potent than morphine, depending on the test system used. It is interesting that compound A given in Scheme 7 has enhanced antitussive activity, whereas its analgesic power is lower than that of morphine. The pharmacology and "modification" of alkaloids is a fascinating and at the same time very important aspect of medicinal chemistry.

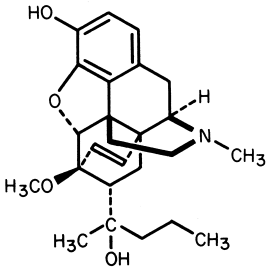
In many cases the source of clinically used alkaloids is still the natural plant (e.g., morphine, quinine), but in some cases industrial synthetic products are preferred (e.g., vincamine, 2,3-dehydroemetine).



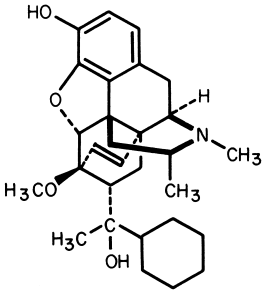
Morphine



Cyclazocine

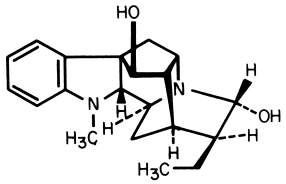


Etorphine

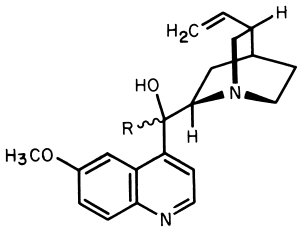


Compound A

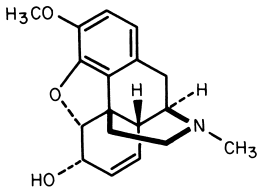
SCHEME 7 Morphine and three of its modifications.



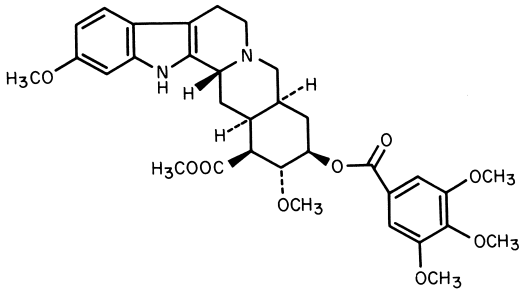
Ajmaline



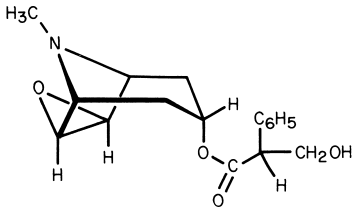
Quinidine, R = ◄H
Quinine, R = ···· *H*



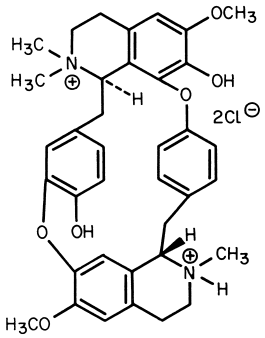
Codeine



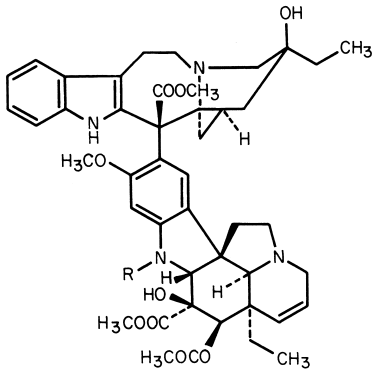
Reserpine



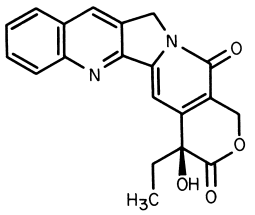
Scopolamine



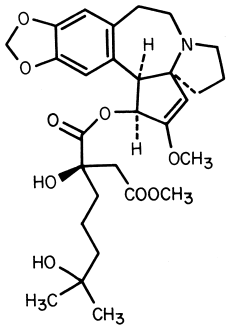
Tubocurarine chloride



Vincaleucoblastine, R = CH_3
Vincristine, R = CHO



Camptothecine



Homoharringtonine

SCHEME 8 Structures of alkaloids in Table IV.

SEE ALSO THE FOLLOWING ARTICLES

ATOMIC SPECTROMETRY • HETEROCYCLIC CHEMISTRY • MICROSCOPY (CHEMISTRY) • ORGANIC CHEMISTRY, COMPOUND DETECTION • ORGANIC CHEMISTRY, SYNTHESIS • PHARMACEUTICALS • X-RAY ANALYSIS

BIBLIOGRAPHY

Atta-ur-Rahman and Basha (1983). "Biosynthesis of Indole Alkaloids," Oxford Univ. Press (Clarendon), London and New York.

Cordell, G. A., ed. (1950–2001). "The Alkaloids: Chemistry and Pharmacology," Vols. 1–55, Academic Press, San Diego.

Cordell, G. A. (1981). "Introduction to Alkaloids: A Biogenetic Approach," Wiley-Interscience, New York.

Dalton, D. R. (1979). "The Alkaloids: The Fundamental Chemistry," Dekker, New York.

Harborne, J. B., and Turner, B. L. (1984). "Plant Chemosystematics," Academic Press, New York.

Hesse, M. (1981). "Alkaloid Chemistry," Wiley, New York.

Mothes, K., and Schütte, H. R. (1969). "Biosynthese der Alkaloide," Deutscher Verlag Wissenschaften, Berlin.

Pelletier, S. W., ed. (1983–1999). "Alkaloids: Chemical and Biological Perspectives," Vols. 1–14, Pergamon, Elmsford, NY.

Specialist Periodical Report (1971–1983). "Alkaloids," Vols. 1–13, The Chemical Society, London.

Waller, G. R., and Nowacki, E. K. (1978). "Alkaloid Biology and Metabolism in Plants," Plenum, New York.

Wexler, P., ed. (1998). "Encyclopedia of Toxicology," Academic Press, San Diego.

Bioconjugate Chemistry

Claude F. Meares
University of California, Davis

GLOSSARY

Antibody A large protein molecule that binds reversibly and with high specificity to a small molecule or a specific region (epitope) of a macromolecule.

Bioconjugate Generally, a synthetic substance in which a functional biological molecule is conjugated (linked) to another substance with useful properties. The latter may be another biological molecule, a synthetic polymer, a small molecule, or even a macroscopic particle or a surface.

Enzyme A protein (or nucleic acid) that catalyzes a chemical reaction, such as hydrolysis of a peptide bond.

Epitope A molecular group, part of a larger molecule, recognized by an antibody or other binding macromolecule. Typically, a small number of amino acid residues in a particular geometric arrangement on the surface of a protein.

Hapten A small molecule recognized by an antibody or other binding macromolecule. Sometimes referred to as a ligand. A hapten may be conjugated to another molecule. For example, biotin binds strongly to the proteins avidin and streptavidin. Biotin may be conjugated to other molecules (see Fig.1), which can then be trapped by binding to immobilized avidin or streptavidin.

Liposome A particle composed of multiple phospholipid molecules, arranged in a bilayer structure to enclose a small aqueous space (see Fig. 2).

ODN An oligodeoxynucleotide, a short, synthetic fragment of deoxyribonucleic acid (DNA).

PEG Poly(ethylene glycol), also called PEO, poly(ethylene oxide) (see Fig. 3). The degree of polymerization **n** is usually specified according to the average number of CH_2—CH_2 units in a polymer molecule.

BIOCONJUGATE CHEMISTRY involves the joining of two molecular functions by chemical or biological means. This includes, among other topics, the conjugation of antibodies, nucleic acids, lipids, carbohydrates, or other biologically active molecules with each other or with other

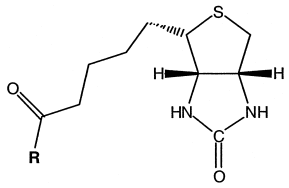


FIGURE 1 D-Biotin conjugated to another molecule, represented by R.

substances that add useful properties (polymers, drugs, radionuclides, toxins, fluorophores, photoprobes, inhibitors, enzymes, haptens, ligands, surfaces, metal particles, etc.).

I. CHEMICAL PRINCIPLES

A. Nucleophiles and Electrophiles

A basic feature of almost all applications of bioconjugate chemistry is the attachment of one chemical species to another. Most often this is done by forming an ordinary covalent chemical bond, taking advantage of naturally occurring reactive groups where possible. For example, proteins commonly have an abundance of nucleophiles (electron-rich atoms or molecular groups) on their surfaces (see Fig. 4), which can react with electrophiles on synthetic reagents.

B. Chemistry of Functional Groups and Reagents

1. Thiols

The thiol (—SH) group, which occurs on the side chain of the amino acid cysteine, is a particularly reactive nucleophile. When treated with an appropriate electrophile, it is not unusual that a single thiol will serve as by far the most

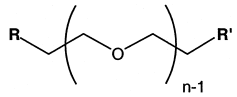


FIGURE 3 PEG. R and R′ refer to attached groups.

reactive site in a protein containing many other, nitrogen- or oxygen-based, nucleophiles. Because of their affinity for metals, thiol groups are also used to attach molecular linkers to metal particles or surfaces.

2. Amines

The amino ($—NH_2$) group, which occurs on the side chain of the amino acid lysine, is also a reactive nucleophile. Because lysine is a common amino acid (a typical antibody molecule of type IgG has more than 80 lysine residues), it is often the target of chemical modification by electrophiles.

3. Disulfides

The disulfide bond (—S—S—), formed by mild oxidation of two thiols, plays a significant role because it is easily broken by mild reducing agents. The disulfide bond is not particularly reactive to nucleophiles or electrophiles, providing a strategy to protect a thiol from reaction. A disulfide bond can also be used to temporarily attach one component of a bioconjugate to the other; reductive cleavage can easily reverse the attachment.

4. Active Esters

a. Carboxylic. Of the formula R—CO—X, where X is a leaving group that is easily displaced, these reactive groups are commonly used to acylate amino groups to

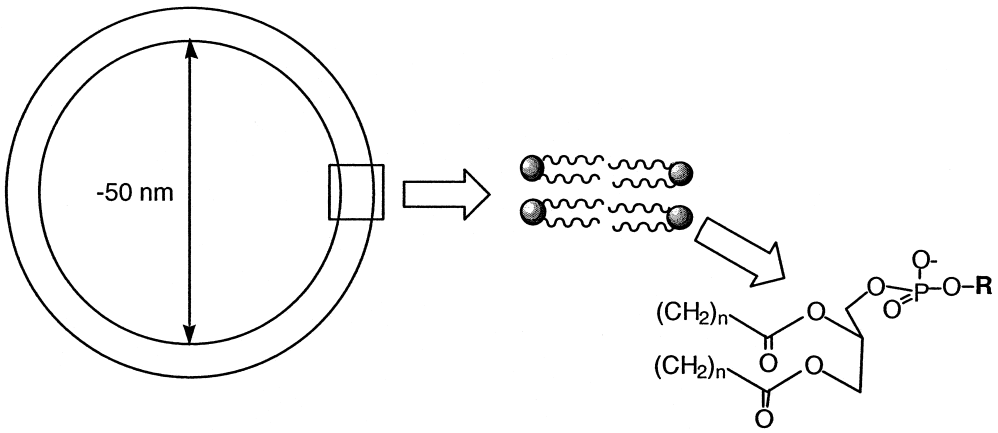


FIGURE 2 Example of a liposome (phospholipid vesicle) conjugated to functional molecules R.

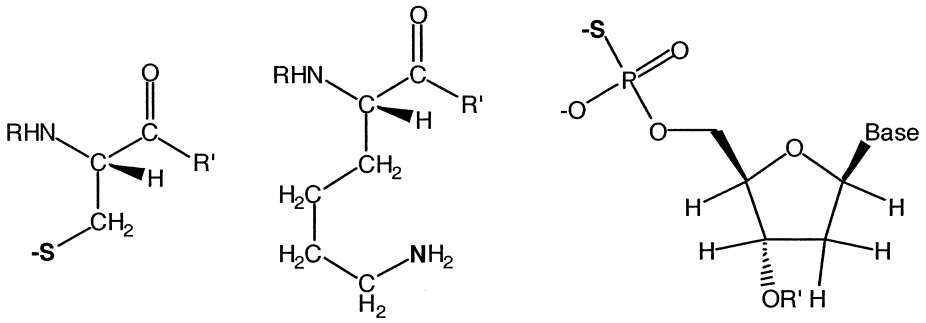


FIGURE 4 Some biological nucleophiles commonly used in bioconjugate chemistry: cysteine, lysine, and a 5′ phosphorothioate group from a synthetic nucleic acid. Nucleophilic atoms are shown in boldface.

form amides R─CO─NHR′, attaching molecular groups R to lysine residues on proteins R′.

b. Imidic. Of the formula $R{-}CNH_2^+{-}X$, used similarly.

5. Isothiocyanates

Of the formula R─N═C═S, these reactive groups are commonly used to modify amino groups to form thioureas R─NH─CS─NHR′, attaching molecular groups R to lysine residues on proteins R′.

6. Alkylating Agents

a. Alkyl halides. Of the formula $R{-}CO{-}CH_2{-}X$, commonly used to alkylate thiol groups to form thioethers $R{-}CO{-}CH_2{-}SR'$, attaching molecular groups R to cysteine residues on proteins R′.

b. Alkenes. Containing a C═C double bond near an activating group, such as carbonyl, used similarly.

7. Photoprobes

Containing photosensitive groups such as aryl azide $R{-}N_3$, diazo $-CN_2-$, or ketones R─CO─R′, which upon irradiation with ultraviolet light produce energetic reagents for covalent attachment to targets.

II. MULTIPLE SITES OF CHEMICAL CONJUGATION: MATHEMATICS OF RANDOM LABELING

Biological molecules such as proteins or nucleic acids can have many sites with similar chemical reactivity, such as lysine residues or phosphate residues. Random copolymers can also contain multiple reactive sites with similar properties. Chemical conjugation involving these sites leads to complex mixtures of products.

Consider a macromolecule with three *equally reactive* sites ($n = 3$) that can be modified by a reagent. We will use a simple 3-bit notation to label each possible product in the resulting mixture (see Fig. 5).

If modification of one site on the macromolecule has no effect on modification of the others (the usual approximation), we can define a *conversion ratio* s as a ratio of concentrations

$$s = \frac{[100]}{[000]} = \frac{[010]}{[000]} = \frac{[001]}{[000]}.$$

(Here we compare each singly modified species to unmodified.)

Using the same assumptions, we discover that doubly modified macromolecules are related to singly modified ones by the same factor s, and to unmodified ones by s^2

$$\frac{[110]}{[100]} = s \quad \frac{[110]}{[000]} = \frac{[110]}{[100]} \bullet \frac{[100]}{[000]} = s \bullet s = s^2.$$

(Note there are three of these.)

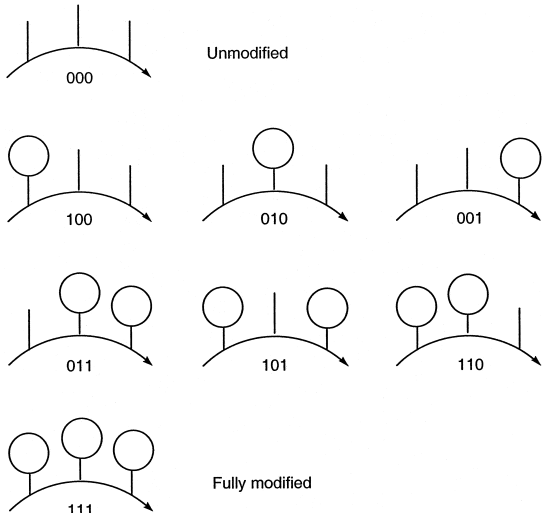


FIGURE 5 Illustration of species produced by random labeling.

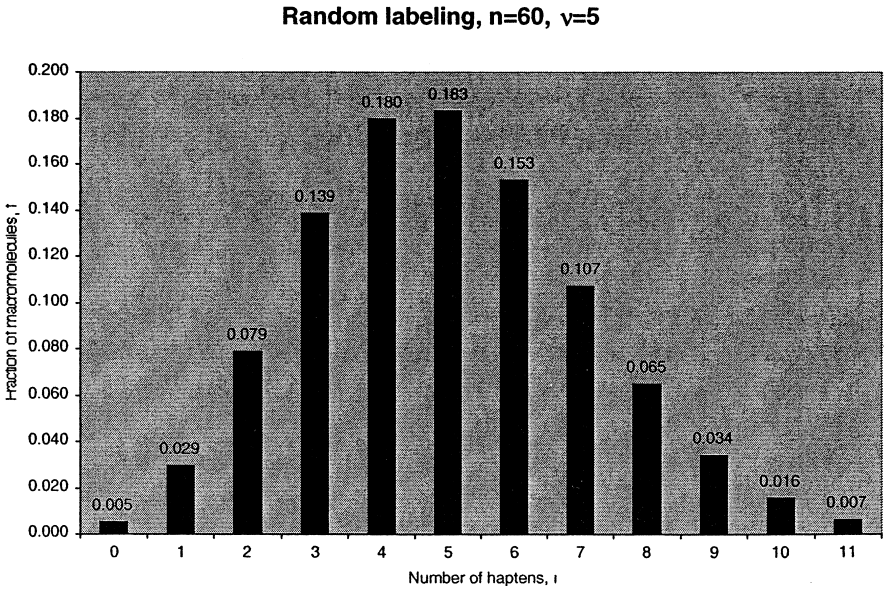


FIGURE 6 Species resulting from random labeling of a macromolecule having $n = 60$ equally reactive sites, such that the average molecule has $n = 5$ haptens attached, according to Eq. (2).

Similarly, for triply modified macromolecules

$$\frac{[111]}{[110]} = s$$

and

$$\frac{[111]}{[000]} = s^3.$$

Another important factor is that there is more than one way to produce a macromolecule with one site modified (e.g., there are three distinct singly-modified species in this example). In general, if a macromolecule has n sites and exactly r of them are randomly modified, then there will be $n!/(n-r)! \bullet r!$ distinct isomers as products.

The *mole fraction* of any of the species is given by, e.g., $X_{001} = [001]/\sum[SSS]$, where

$$\sum[SSS] = [000]\left(\sum \frac{[SSS]}{[000]}\right)$$
$$= [000]\,(I + s + s + s + s^2 + s^2 + s^2 + s^3)$$
$$= [000] \sum_{r=0}^{n} \frac{n!}{(n-r)! \bullet r!} s^r = [000]Z$$

Note that Z is easy to calculate if we know s and n.

So the mole fraction of any species is, e.g., $X_{001} = [001]/\sum[SSS] = [001]/\{[000]Z\} = s/Z = s/(1 + 3s + 3s^2 + s^3)$. In general, the mole fraction f_r of product with a particular degree of modification r is (Eq. 1):

$$f_r = \frac{\dfrac{n!}{(n-r)! \bullet r!} s^r}{\sum_k \dfrac{n!}{(n-k)! \bullet k!} s^k}$$

The *average value* of r in a real mixture of modified macromolecules is denoted ν, which is given by Eq. (2):

$$\upsilon = \sum_r f_r \bullet r = \frac{\sum_r \dfrac{n!}{(n-r)! \bullet r!} r \bullet s^r}{\sum_k \dfrac{n!}{(n-k)! \bullet k!} s^k}$$

Figure 6 plots the fraction f_r of macromolecules having exactly r haptens attached, up to the value $r = 11$. Note that each column contains contributions from $n!/(n-r)! \bullet r!$ distinct isomers. Eq. (2) can be used to calculate f_r for larger values of r.

III. MOLECULAR CLONING: MODERN PREPARATIVE METHODOLOGY

Gene fusion, whereby the gene coding for one protein is connected to the gene coding for another, is one method by which site-specific attachment of two proteins can be achieved. Site-directed mutagenesis, in which one amino acid residue of a protein is genetically altered to produce a mutant protein containing a single cysteine residue, is suited for site-specific conjugation of other molecules or site-specific immobilization of the protein onto a surface.

IV. ANALYSIS OF BIOCONJUGATES

A. Methods

Characterization of bioconjugates frequently relies on assays of the properties of the components (enzyme

activity, antibody binding, radioactivity, fluorescence, etc.). In addition, gel electrophoresis, high-performance liquid chromatography, and mass spectrometry are frequently employed to confirm composition, molecular weight, homogeneity, and even details regarding sites of attachment. Bioconjugates involving larger components such as particles or surfaces are characterized using microscopy and other surface-science techniques.

B. Challenges

The precise specification of the chemical linkages formed when a protein or other biomolecule is chemically attached to another biomolecule, the product often having molecular weight >100,000 daltons and multiple possible points of attachment, is rarely attempted. In the past and still today, researchers often work with mixtures of molecular conjugates having different points of attachment but similar composition. In special cases, a particularly reactive naturally occurring residue such as cysteine permits easy preparation of homogeneous products.

V. NOMENCLATURE

The complexity of typical bioconjugates has so far defeated efforts to systematize their nomenclature. It is customary to name products by referring to starting materials; thus antibody–enzyme conjugates are named according to the individual molecules involved.

VI. EXAMPLES

A. Targeted Therapy

An important application of bioconjugates involves the treatment of disease by specifically targeting disease sites. Combining the properties of an antibody, synthetic polymer, or liposome with those of a drug, nucleic acid, or radionuclide is currently an important research activity. Some therapeutic products are now commercially available, and many clinical trials are under way.

The important properties of antibodies are their specific binding to molecular targets on diseased cells and the physiological consequences of their large size, which influence where they localize and how long they remain there. Artificial polymers such as PEG share some of the latter properties, as do liposomes. Smaller molecules can be used in place of these targeting moieties, including smaller fragments of antibodies that retain binding specificity (Fab fragments, single-chain antigen binding proteins, targeting peptides, etc.) and non-protein molecules based on nucleic acids (aptamers) or other structures found to bind to or accumulate in biological targets. A typical use of the polymer PEG is to modify the biological properties of proteins. By attaching PEG groups to lysine side chains, it is possible to reduce the likelihood that the immune system will mount a response to the protein. This permits repeated use of the protein over several cycles of treatment. Another application of PEG conjugation is to modify the rate at which a protein leaves the circulation, allowing therapeutic bioconjugates to circulate for longer periods. The important properties of nucleic acids and their analogs involve modification of gene expression processes in target cells. This may occur by the insertion of new genes into cells, or by interference with the expression of existing genes by "antisense" binding. Antisense oligodeoxyribonucleotides (ODNs) can inhibit gene expression by hybridization to complementary messenger RNA sequences inside cells.

B. Imaging

Bioconjugates of low molecular weight are used extensively in imaging sites of human disease by nuclear techniques such as ^{99m}Tc scintigraphy (making images from gamma photons). Here a coordination complex formed between the metal technetium and an organic ligand with targeting properties (e.g., a small peptide, a hormone analog, etc.) serves to localize the radiotracer in the target tissue. Positron Emission Tomography (PET) with ^{18}F-labeled fluorodeoxyglucose or other small molecules provides higher resolution images by detecting two annihilation photons in coincidence.

Magnetic Resonance Imaging (MRI) sometimes makes use of bioconjugates containing paramagnetic metals such as gadolinium or manganese conjugated to polymers to modify their relaxivity properties.

Optical imaging probes that fluoresce in the near-infrared wavelength region can be conjugated with targeting molecules and used for in vivo imaging.

1. Separation and Analysis *in vitro*

Immobilization of biomolecules by linking them to surfaces, such as the wells of microtiter plates, magnetic beads, or chromatography supports for subsequent, specific binding of target molecules is often the first step to *in vitro* analysis. The nature of the linkage may be as simple as adsorption of a protein to a hydrophobic surface, or it may involve covalent attachment of the biomolecule to reactive groups on the surface using the chemical principles discussed above. With modern techniques, it is usually possible to accomplish this surface attachment without significant loss of biological properties.

Conjugates containing antibodies that specifically bind to a molecule of interest are widely used in bioanalytical

chemistry. Conjugates of antibodies with enzymes are used for assays of all sorts of biological molecules by enzyme-linked immunosorbent assay (ELISA). For example, a protein in a biological fluid can be analyzed by first trapping it using an immobilized antibody that binds specifically and strongly to an epitope on that particular protein. After washing away unbound substances, a second antibody that binds another epitope on the target protein is allowed to bind to the immobilized target. This second antibody is conjugated to an enzyme; at this point, the sample contains a number of immobilized enzyme molecules, related to the number of target protein molecules in the original biological fluid. The immobilized enzyme molecules are then quantified by standard techniques of enzymology, such as catalytic production of a light-absorbing product that can be measured spectrophotometrically. Western blotting is another application of antibody conjugates, wherein electrophoretic separation of proteins is followed by blotting the separated bands onto a special membrane. The bands containing a specific antigen are then stained and visualized by procedures not unlike those used for ELISA, except that the colored product is precipitated where the bands are.

2. Biophysical Studies

Bioconjugate chemistry plays a prominent role in studying conformational changes in macromolecules or the binding of one macromolecule to another. Techniques such as chemical cross-linking, fluorescence energy transfer, fluorescence polarization, electron paramagnetic resonance, or various other spectroscopy or microscopy experiments, generally involve preparation of a bioconjugate tagged in some way by a reporter molecule. Fluorescence energy transfer from an excited donor moiety to an acceptor moiety depends on the distance between donor and acceptor; it is useful for measuring distances in the 10–100 Å range, appropriate for macromolecular complexes. Fluorescence polarization is sensitive to difference in the rotational diffusion (tumbling) of molecules in solution, and can readily reveal the binding of a small molecule to a macromolecule. Electron paramagnetic resonance provides related information about changes in tumbling behavior, and can also be used to study the distances between paramagnetic centers. Electron microscopy can show the spatial relations between metal clusters attached to macromolecules as probes. Recently, an approach has emerged that involves cleaving the sugar-phosphate backbone of a nucleic acid or the polypeptide backbone of a target protein, using a cutting protein that binds to the target. The cutting protein is prepared by conjugation with a small chemical reagent, which acts as an artificial nuclease or protease. Using tools such as nucleic acid sequencing or Western blotting, the sites on the target macromolecule that are within reach of the cutting reagent can be identified. Chemical reagents designed for use in all these biophysical studies are commercially available from specialty vendors.

SEE ALSO THE FOLLOWING ARTICLES

BIOPOLYMERS • CARBOHYDRATES • ENZYME MECHANISMS • NUCLEIC ACID SYNTHESIS • PROTEIN SYNTHESIS

BIBLIOGRAPHY

Bioconjugate Chemistry, a journal published by the American Chemical Society (*http://pubs.acs.org/BC*).

Hermanson, G. T. (1996). "Bioconjugate Techniques," Academic Press, San Diego, CA.

Lundblad, R. L. (1995). "Techniques in Protein Modification," CRC Press, Boca Raton, FL.

Means, G. E., and Feeney, R. E. (1971). "Chemical Modification of Proteins," Holden-Day, San Francisco, CA.

Aslam, M., and Dent, A. (1998). "Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences," Grove's Dictionaries Inc., New York.

"Pierce Catalog and Handbook" (1994, 1995). Life Science & Analytical Research Products, Pierce Chemical Company, Rockford, IL.

Wong, S. S. (1991). "Chemistry of Protein Conjugation and Cross-Linking," CRC Press, Inc., Boca Raton, FL.

Carbohydrates

Hassan S. El Khadem
The American University

GLOSSARY

Aldoses Chiral polyhydroxyalkanals having three or more carbon atoms (polyhydroxyalkanals having fewer than three carbons are achiral). Aldoses having enough carbon atoms to form five- or six-membered oxygenated rings readily undergo cyclization by intramolecular nucleophilic attack of suitably located hydroxyl groups onto the aldehydic groups. Saccharides possessing five-membered rings are called furanoses, and those possessing six-membered ones, pyranoses.

Anomeric configuration The cyclization of an aldose or a ulose to form a furanose or a pyranose ring occurs by intramolecular nucleophilic attack of a suitably located hydroxyl group on a carbonyl function (present in the same molecule). This cyclization results in the formation of a chiral hemiacetal group in place of an achiral carbonyl group, which increases the number of chiral centers present in the molecule by one. The configuration of the newly created center is designated by the greek letter α or β, followed by the D or L notation of the chiral center farthest from the carbonyl. Thus, one speaks of α-D-aldopento-furanoses or β-L-hexopyranuloses.

Furanoses Aldoses having four carbon atoms or more and uloses having five carbon atoms or more can form five-membered rings. Such cyclic monosaccharides are called furanoses because they are related to tetrahydrofuran. Aldopentoses having such five-membered rings are designated aldopentofuranoses, and hexuloses, hexofuranuloses. In solution, furanoses exist mainly in envelope and twist conformations, alternating in a wavelike motion between them.

Monosaccharides Monomeric saccharides, which are grouped according to the number of carbon atoms in their skeleton into pentoses, hexoses, and so on. Members of these groups may exist in the form of polyhydroxy aldehydes, called aldoses, or polyhydroxy ketones, called uloses. Accordingly, certain monosaccharides are designated aldopentoses, other hexuloses, and so on.

Mutarotation Designates changes in optical rotations observed over a period of time. The phenomenon occurs when chiral substances equilibrate in solution with one or more isomeric forms. Saccharides usually crystallize in structurally pure form; for example, D-glucose usually separates in the α-D-glucopyranose form. When this is dissolved in water, it undergoes

equilibration to yield a complex mixture of six isomers (two furanoses, two pyranoses, an acyclic carbonyl form, and a hydrated carbonyl form). The rotation of the equilibrium mixture is naturally different from that of the original pure α-D-glucopyranose, which accounts for the observed change in rotation.

Oligosaccharides Low molecular weight polymers made up of 2 to 10 monosaccharide units linked together by acetal linkages. Oligosaccharides are classified according to degree of polymerization into disaccharides, trisaccharides, and so on. They are also grouped into reducing and nonreducing oligosaccharides. If one of the terminal monosaccharides possesses a hemiacetal group, the oligomer is referred to as a reducing oligosaccharide (since it is susceptible to mild oxidants). If, on the other hand, it does not possess a hemiacetal group (such as when two anomeric hydroxyl groups on adjacent monosaccharides are involved in acetal bond formation), the oligosaccharide is referred to as nonreducing. Reducing oligosaccharides mutarotate, because in solution one of their terminal monosaccharides is in equilibrium with other forms, while nonreducing oligosaccharides do not mutarotate.

Polysaccharides Polymers composed of more than 10 monosaccharide units linked by acetal bridges. Most naturally occurring polysaccharides have degrees of polymerization ranging between 10^2 and 10^5. Linear polysaccharides such as cellulose are usually microcrystalline, are insoluble in water, and form strong films and fibers, whereas highly branched polysaccharides such as gums form gets and produce brittle films.

Pyranoses Aldoses having five carbon atoms or more and uloses having six carbon atoms or more can form strainless six-membered rings. Such cyclic forms of saccharides are called pyranoses, because they are related to tetrahydropyran. Aldopentoses having six-membered rings are called aldopentopyranoses, and hexuloses possessing such rings are designated hexopyranuloses. The most stable conformations of pyranoses are the two chair conformations $^1C_4$ and $^4C_1$, of which the latter is usually the more stable.

Saccharide (Sugarlike; synonymous with carbohydrate). It is often prefixed by "mono," "oligo," or "poly" to designate degree of polymerization.

Sugars Sweet-tasting saccharides (monosaccharides and disaccharides are sweet tasting, whereas most trisaccharides and all higher saccharides are devoid of taste). Because the term *sugar* encompasses compounds possessing a physiological rather than a chemical property, and because the term has been used indiscriminately to designate different saccharides (sucrose by the public and D-glucose by the medical profession), it is slowly being replaced in chemistry texts by less ambiguous names such as monosaccharides and disaccharides.

Uloses (also known as ketoses) Chiral polyhydroxyalkanones that possess at least four carbon atoms in their skeleton (lower members are achiral). Uloses that possess a sufficient number of carbon atoms to cyclize in strainless rings do so and yield furanuloses and pyranuloses.

CARBOHYDRATES are monomeric, oligomeric, or polymeric forms of polyhydroxyalkanals or polymeric forms of polyhydroxyalkanals or polyhydroxyalkanones, which are collectively called monosaccharides. In the nineteenth century, when the empirical formulas of organic compounds were determined, it was found that the sugars and polysaccharides known at the time had the formula $C_n(H_2O)_y$. *They were accordingly given the name carbohydrates* (i.e., hydrates of carbon). Today's usage of the word *carbohydrate* applies to a large number of monomeric, oligomeric, and polymeric compounds, which do not necessarily have their hydrogen and oxygen atoms in the molecular ratio of 2:1, but which belong to the group of compounds called *monosaccharides* or can be readily converted to members of that group by hydrolysis.

I. INTRODUCTION

A. Historical Background

The origins of carbohydrate chemistry can be traced back to the civilizations of antiquity. For example, the manufacture of beer and wine by alcoholic fermentation of grain starch and grape sugar is well documented on the walls of ancient Egyptian tombs. The isolation of cellulose fibers from cotton and flax was started by the civilizations of the Far and Near East and was introduced by the Greeks to Europe. Gums and resins were valued commodities at the beginning of the Christian era.

The isolation of sucrose from the juice of sugarcane marks an important milestone in sugar chemistry. In the Far East, where this plant grew, sugar was isolated as a yellowish syrupy concentrate that crystallized, on standing, into a brown mass, and a number of Chinese recipes dating from the fourth century describe in detail how the sugarcane juice was concentrated (Fig. 1). It is interesting that the Sanskrit word *sugar* means "sweet sand," which aptly describes the properties of crushed, raw sugar. Significant progress in the manufacture of sugar occurred after the French Revolution, when Europe under blockade had to rely on the sugar beet for the manufacture of this important commodity. Improved methods of isolation and

FIGURE 1 Reproduction of a seventh-century Chinese drawing representing the manufacture of sugar.

refining were developed, including treatment of the syrup with lime to precipitate the calcium complex, regeneration of the sugar with sulfur dioxide, and decolorization with animal charcoal.

B. Importance of Carbohydrates

In addition to the manufacture of sucrose and paper, which are today important industries, carbohydrates play a major role in a number of other industries. These include (1) the food industry, which uses huge amounts of starch of various degrees of purity in the manufacture of baked goods and pastas, of gums in food processing, and of mono- and oligosaccharides as sweeteners and employs carbohydrates in fermentation to make beer and wine; (2) the textile industry, which, despite the advent of synthetics, is still dependent to a large extent on cellulose; (3) the pharmaceutical industry, particularly in the areas of antibiotics, intravenous solutions, and vitamin C; and (4) the chemical industry, which produces and markets several pure sugars and their derivatives.

Carbohydrates also play a key role in the process of life; the master molecule DNA is a polymer made up of repeating units composed of four nucleotides of 2-deoxy-D-*erythro*-pento-furanose (2-deoxy-D-ribose), the sequence of which constitutes the coded template responsible for replication and transcription. Saccharide derivatives also form part of many vital enzyme systems, specifically as coenzymes.

The best-known use of carbohydrates is undoubtedly in nutrition, as members of a major food type that are metabolized to produce energy. Although the average percentage of carbohydrates consumed by humans in comparison with other food types differs from country to country, and figures are often inaccurate or unavailable, the percentage for the world as a whole has been estimated to be more than 80%. The exothermic reactions that produce energy in the cell are the outcome of a number of complex enzyme cycles that originate with hexoses and end up with one-, two-, or three-carbon units. The energy released from these reactions is stored in the cell in the form of a key sugar intermediate, adenosine triphosphate (ATP), and released when needed by its conversion to the di- or monophosphate. Finally, carbohydrates are the most abundant organic components of plants (>50% of the dry weight) and therefore constitute the major part of our renewable fuels and the starting material from which most of our fossil fuels were made.

C. Classification

Carbohydrates are classified according to their degree of polymerization into monomeric carbohydrates, which include monosaccharides and their derivatives, and polymeric carbohydrates, which comprise oligosaccharides, polysaccharides, DNA, and RNA. The polymeric carbohydrates differ in the type of bridge that links their monosaccharide units. Thus, oligosaccharides and polysaccharides are polyacetals, linked by acetal oxygen bridges, whereas DNA and RNA are poly(phosphoric) esters, linked by phosphate bridges. In addition to these well-defined groups of carbohydrates, there exist a number of derivatives, for example, antibiotics, that are best studied as a separate group, because some of their members may be monomeric, whereas others are oligomeric.

1. Monosaccharides

Monosaccharides are chiral polyhydroxyalkanals or polyhydroxyalkanones that often exist in cyclic hemiacetal forms. Monosaccharides are divided into two major groups according to whether their acyclic forms possess an aldehyde group or a keto group, that is, into aldoses or ketoses (glyculoses). Each of these, in turn, is classified according to the number of carbon atoms in the monosaccharide chain (usually 3–10) into trioses, tetroses, pentoses, hexoses, and so on. By prefixing "aldo" to these names, one can define more closely a group of aldoses, for example, aldopentoses, whereas for ketoses, it is customary to use the ending "ulose," as in hexulose. Finally monosaccharides can be grouped according to the size of their rings into five-membered furanoses and six-membered pyranoses. It should be noted that, in order to form a furanose ring, four carbon atoms and one oxygen atom are needed, so that only aldotetroses and higher aldoses, and 2-pentuloses and higher ketoses, can cyclize in this ring form. Similarly, in order to form a pyranose ring five

TABLE I Monosaccharides: Aldoses

Monosaccharide	Aldose[a]	No. of chiral carbons	Aldofuranose	Aldopyranose
Triose	Aldotriose	1	—	—
Tetrose	Aldotetrose	2	Tetrofuranose; aldotetraofuranose	—
Pentose	Aldopentose	3	Pentofuranose; aldopentofuranose	Pentopyranose: aldopentopyranose
Hexose	Aldohexose	4	Hexofuranose; aldohexofuranose	Hexopyranose; aldohexopyranose
Heptose	Aldoheptose	5	Heptofuranose; aldoheptofuranose	Heptopyranose; aldoheptopyranose
Octose	Aldooctose	6	Octofuranose; aldooctofuranose	Octopyranose; aldooctopyranose
Nonose	Aldononose	7	Nonofuranose; aldononofuranose	Nonopyranose; aldononopyranose
Decose	Aldodecose	8	Decofuranose: aldodecofuranose	Decopyranose; aldodecopyranose

[a] Although an achiiral aldobiose (glycolic aldehyde) exists, it is not considered to be a saccharide because, by definition, a saccharide must contain at least one asymmetric carbon atom.

carbon atoms and one oxygen atom are required, so that only aldopentoses and 2-hexuloses, as well as their higher analogs, can cyclize in this form. Ultimately, by combining the ring type to the names used above (e.g., aldopentose or hexulose), such combination names as aldopentofuranose and hexopyranulose can be formed, which define without ambiguity the group to which a monosaccharide belongs (see Table I and II).

As their name denotes, monosaccharides are monomeric in nature and, unlike the oligosaccharides and polysaccharides, which will be discussed later (Sections III and IV), they cannot be depolymerized by hydrolysis to simpler sugars. Monosaccharides and oligosaccharides are soluble in water; their solutions in water are often sweet tasting, and this is why they are referred to as sugars.

2. Oligosaccharides

Oligosaccharides and polysaccharides are polyacetals that respectively have, as their names denote (*oligo* = "few"; *poly* = "many," in Greek), a low (2–10) or a high (>10) degree of polymerization (DP). They are composed of a number of monosaccharides linked together by acetal oxygen bridges and yield on depolymerization (hydrolysis) one or more types of monosaccharide. Oligosaccharides are further grouped into (1) simple (true) oligosaccharides which are oligomers of monosaccharides that yield on *complete* hydrolysis *only* monosaccharides; and (2) conjugate oligosaccharides, which are oligomers of monosaccharides linked to a nonsaccharide, such as a lipid aglycon, usually a long chain fatty acid, alcohol or amine, or a carbocyclic steroid or terpenoid. The carbohydrate portion of these compounds is resposible for their specificity and is involved in cell recognition. Two main types of compounds are recognized: *glycolipids*, which can be of animal, plant, or microbial origin, and *gangliosides*, which contain sialic acid and are found in large amounts in cerebral tissues of patient defficient in the enzyme *N-acetylhexosaminidase*. There are two ways to classify simple oligosaccharides further. The first is, according to DP, into disaccharides, trisaccharides, tetrasaccharides, and so on, and the second, according to whether the oligomer chain has at one end a hemiacetal or a hemiketal function (a latent aldehyde or keto group). Such terminal groups, if present, are readily converted to carboxylic groups by mild oxidants, and accordingly, oligosaccharides

TABLE II Monosaccharides: Ketoses (Glyculoses)

Monosaccharide	Ketose[a]	No. of chiral carbons	Ketofuranose	Ketopyranose
Tetrose	Tetrulose	1	—	—
Pentose	Pentulose	2	Pentulofuranose	—
Hexose	Hexulose	3	Hexulofuranose	Hexulopyranose
Heptose	Heptulose	4	Heptulofuranose	Heptulopyranose
Octose	Octulose	5	Octulofuranose	Octulopyranose
Nonose	Nonulose	6	Nonulofuranose	Nonulopyranose
Decose	Deculose	7	Deculofuranose	Deculopyranose

[a] Although an achiiral triulose (1,3-dihydroxyacetone) exists, it is not considered to be a saccharide because it lacks an asymmetric carbon atom.

TABLE III Oligosaccharides[a]

Reducing		Nonreducing	
Homooligosaccharides	**Heterooligosaccharides**	**Homooligosaccharides**	**Heterooligosaccharides**
Simple oligosaccharides			
Disaccharides			
Maltose	Lactose	Trehalose	Sucrose
4-α-D-Glc$p$-D-Glc	4-β-D-Gal$p$-α-D-Glc	α-D-Glc$p$-α-D-Glc$p$	β-D-Fru$f$-α-D-Glc$p$
Cellobiose	Lactulose	Isotrehalose	Isosucrose
4-β-D-Glc$p$-D-Glc	4-β-D-Gal$p$-D-Fru	β-D-Glc$p$-β-D-Glc$p$	α-D-Fru$f$-β-D-Glc$p$
Isomaltose	Melibiose		
6-α-D-Glc$p$-D-Glc	6-α-D-Gal$p$-D-Glc		
Gentiobiose	Turanose		
6-β-D-Glc$p$-D-Glc	3-α-D-Glc$p$-D-Fru		
Trisaccharides			
Maltotriose	Manninotriose		Raffinose
4-α-D-Glc$p$-Maltose	6-α-D-Gal$p$-Melibiose		6-α-D-Gal$p$-Sucrose
Tetrasaccharides			
Maltotetraose			Stachyose
4-α-D-Glc$p$-Maltotriose			6-α-D-Gal$p$-Raffinose
Pentasaccharides			
Hexasaccharides			
Heptasaccharides			
Octasaccharides			
Nonasaccharides			
Decasaccharides			
Conjugate oligosaccharides			Glycolipids
			Gangliosides

[a] Glc, Glucose; Gal, galactose; Fru, fructose; f, furanose; p, pyranose.

possessing these groups are referred to as *reducing*, in contradistinction to those that resist such oxidation and are designated *nonreducing* (see Table III). Thus. whereas all monosaccharides are reducing, there are reducing and nonreducing disaccharides, trisaccharides, and so on. Because monosaccharides and simple oligosaccharides are soluble in water and sweet tasting, they are called sugars. It should be noted, however, that the sweetness of oligosaccharides decreases with increasing DP and disappears at DP 4–5.

3. Polysaccharides

Polysaccharides and oligosaccharides are polymeric in nature and are structurally similar (both are polyacetals having oxygen bridges linking the monosaccharide monomers), but they may differ markedly in DP; the polysaccharides may reach a DP of 10^5, whereas, by definition, the maximum DP for oligosaccharides is 10. Although, by convention, compounds having a DP of 11 or more are designated polysaccharides, a difference between the properties of a lower polysaccharide with a DP of 11 and those of a higher (DP = 10) oligosaccharide can hardly be detected. However, most polysaccharides have a much higher DP than oligosaccharides, which renders quite significant the sum of the gradual changes that occur in their physical properties with increasing DP. For example, because, with a rise in DP, the solubility decreases and the viscosity increases, some higher polysaccharides, for example, cellulose, are completely insoluble in water (all oligosaccharides are soluble), while the increase in viscosity may cause the solutions of other polysaccharides to set and gel.

There are several ways to classify polysaccharides; a common one is to group them according to their sources, that is, into plant and animal polysaccharides, and then subdivide the former into skeletal polysaccharides (cellulose, etc.), reserve polysaccharides (starch, etc.), gums and mucilages, algal polysaccharides, bacterial polysaccharides, and so on. The disadvantage of this classification is that it tells us very little about the chemistry of these polymers.

The classification used in chemistry texts usually distinguishes between (1) simple (true) polysaccharides that

TABLE IV Polysaccharides

Homopolysaccharides		Heteropolysaccharides	
Linear	**Branched**	**Linear**	**Branched**
Simple polysaccharides			
Amylose (α-D-glucan)	Amylopectin	Mannans	Gums
Cellulose (β-D-glucan)	Glycogen	Xylans	Mucilages
Chitin (D-glucosaminan)			Pectins
			Algin
			Agar
			Bacterial polysaccharides
Conjugate polysaccharides			Peptidoglycans
			Glycoproteins
			Lectins

afford on depolymerization only mono- and oligosaccharides or their derivatives (esters or ethers) and (2) conjugate polymers made up of a polysaccharide linked to another polymer, such as a peptide or a protein (to form a glycopeptide or a glycoprotein). True polysaccharides are, in turn, grouped into two major classes: (1) homopolysaccharides, which are polymers having as a repeating unit (monomer) one type of monosaccharide; and (2) heteropolysaccharides, which are made up of more than one type of monosaccharide. Because the shape of polymers significantly influences their physical properties, each of these types of polymer is further divided into linear and branched polysaccharides (Table IV).

4. DNA, RNA, Nucleotides, and Nucleosides

Unlike oligo- and polysaccharides, which are polyacetals linked by oxygen bridges, DNA and RNA are polyesters linked by phosphate bridges. DNA is the largest known polymer; its DP exceeds 10^{12} in human genes and decreases as the evolution ladder is descended. This giant molecule plays a key role in replication and in transcription. It achieves the latter by doubling one of its strands with a smaller polymer, mRNA, which, in turn, binds with a string of oligomers, tRNA, to form the peptide chain. The monomers of both DNA and RNA are made up of phosphorylated 2-deoxy-D-*erythro*-pentofuranosyl- and D-ribofuranosyl-purine and -pyrimidine bases, designated nucleotides. The latter can undergo hydrolysis of their phosphoric ester groups to afford simpler monomers, the nucleosides. Thus it is apparent that this group can also be divided according to DP into monomers (nucleosides and nucleotides), oligomers (tRNA), and polymers (DNA and mRNA).

Some carbohydrate derivatives may belong to more than one of the above-mentioned groups. For example, carbohydrate-containing antibiotics may be monosaccharide or oligosaccharide in nature.

II. MONOSACCHARIDES

Monosaccharides (monomeric sugars) exist as chiral polyhydroxyalkanals, called aldoses, or chiral polyhydroxyalkanones, called ketoses, which are further classified (see Tables I and II) according to the number of carbon atoms in their chains into trioses, tetroses, pentoses, hexoses, and so on, and according to the type of ring they form into furanoses and pyranoses (five- and six-membered rings).

A. Structure, Configuration, and Conformation of Monosaccharides

1. Structure

The presence of an aldehyde group in aldoses such as D-glucose was established by the fact that (1) aldoses react with carbonyl-group reagents, giving oximes and hydrazones; and (2) aldoses are oxidized to acids that possess the same number of carbon atoms. Furthermore, by estimating the number of acetyl groups in fully acetylated aldoses, it is possible to determine the number of hydroxyl groups present in the starting aldoses (e.g., five OH groups in aldohexoses) and to determine their structural formulas.

2. Configuration

The system commonly used to represent three-dimensional linear molecules such as monosaccharides two dimensionally is the Fischer projection formula. This affords an unambiguous way of depicting monosaccharides, as follows. (1) The carbon chain is drawn vertically,

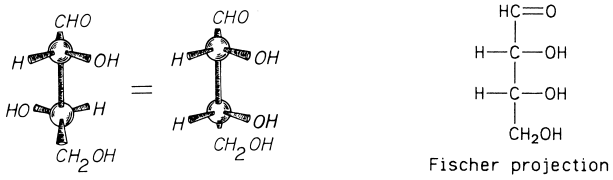


FIGURE 2 Models and formula of D-Erythrose. The two models on the left are rotamers of the compound on the right.

with the carbonyl group at (or nearest to) the top and the last carbon atom in the chain (i.e., the one farthest from the carbonyl group) at the bottom. (2) Each carbon atom is rotated around its vertical axis until all of the vertical (C—C) bonds in the chain lie below an imaginary curved plane such as that of a rolled piece of paper, and all of the horizontal bonds (parallel to the *x* axis) lie above the plane of the paper. The curved plane is then flattened, and the projection of the molecule is represented as viewed (see Fig. 2).

a. Relative and absolute configuration. The relative configuration of D-glucose was established by Emil Fischer in 1891 and constituted at the time a monumental achievement, for which he earned a Nobel prize. Nowadays, the determination of the absolute configuration of a monosaccharide offers no difficulty, because the configurations of a large number of related compounds are available. The unknown is simply converted to a compound of known configuration by means of reactions that do not affect the configuration at the chiral center(s).

Since D-glucose is an aldohexose, it must possess four chiral carbon atoms and can exist in $2^4 = 16$ stereoisomers (see Table I). In order to determine which of these isomers is actually D-glucose, a reference compound of known absolute configuration is needed, which can be prepared from, or converted to, D-glucose. Because the first determination of the absolute configuration of an organic compound had to await the advent of X-ray crystallographic techniques developed in the middle of the twentieth century, such a reference compound did not exist in Fischer's time. This is why he was able only to propose a relative configuration for the monosaccharides known in his time. He chose, as the reference compound, the dextrorotatory form of glyceraldehyde, now designated D-(+)-glyceraldehyde, and arbitarily assumed that the OH attached to C-2 in this compound is to the right when represented by a Fischer projection formula (see Fig. 3).

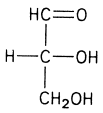


FIGURE 3 Fischer projection formula of D-glycer-aldehyde.

It was fortunate that the arbitary assignment was later confirmed by X-ray crystallography. Otherwise, all of the configurations assigned to carbohydrates during the 60 years that followed Fischer's assignment would have been in error, causing confusion and chaos in the chemical literature.

Once the absolute configuration at C-2 of the D-(+) isomer of glyceraldehyde had been established, it could be related to the configuration of the corresponding center in any aldose that is obtained from this isomer by ascending the series or that yields this isomer by repeated descending degradations (Fig. 4). It should be stated at this point that, because C-2 in glyceraldehyde corresponds to the chiral center farthest from the carbonyl group in any aldose or ketose, it is possible to group monosaccharides into two families: one related to D-(+)-glyceraldehyde and designated by the prefix D, and one related to L-(−)-glyceraldehyde and designated by the prefix L. Monosaccharides of the D family all have the (*R*) configuration at the chiral center farthest from the carbonyl (the OH attached to this chiral carbon is to the right in a Fischer projection). Conversely, monosaccharides of the L family have the (*S*) configuration at this center (the OH group is to left).

b. Configuration of the acyclic form of aldoses. Fischer used an ingenious method to determine whether in an aldose molecule, the configurations of the chiral centers that are equidistant from the center (e.g., C-2 and C-4 in a pentose) had the same sign. He determined whether conversion of the two groups situated at the "top" and the "bottom" of an aldose molecule (CHO and CH_2OH) into the same type of group, for example, a carboxylic group, rendered the product achiral (afforded a meso compound). If so, he could conclude that a plane of symmetry was created during the conversion (oxidation) and that the configuration of the chiral centers situated at equal distances from this plane of symmetry was identical. He also reasoned that, if the dicarboxyclic acid possessed an axis of symmetry, it could be obtained from only one aldose, whereas if it lacked an axis, it could be obtained from two different aldoses, an aldose having the aldehyde group at the top and the primary hydroxyl group at the bottom, and another aldose having the aldehyde group at the bottom and the primary hydroxyl group at the top. Using this reasoning. Fischer was able to determine the configuration of all the aldoses known in his time.

c. Cyclic structures and anomeric configuration. Soon after the configuration of the acyclic form of glucose and the other aldoses had been established, it became apparent that these structures could not represent the major components of the equilibrium mixture. Thus, the IR spectrum of D-glucose does not exhibit a strong carbonyl band at 1700 cm^{-1}, and its ^1H NMR spectrum

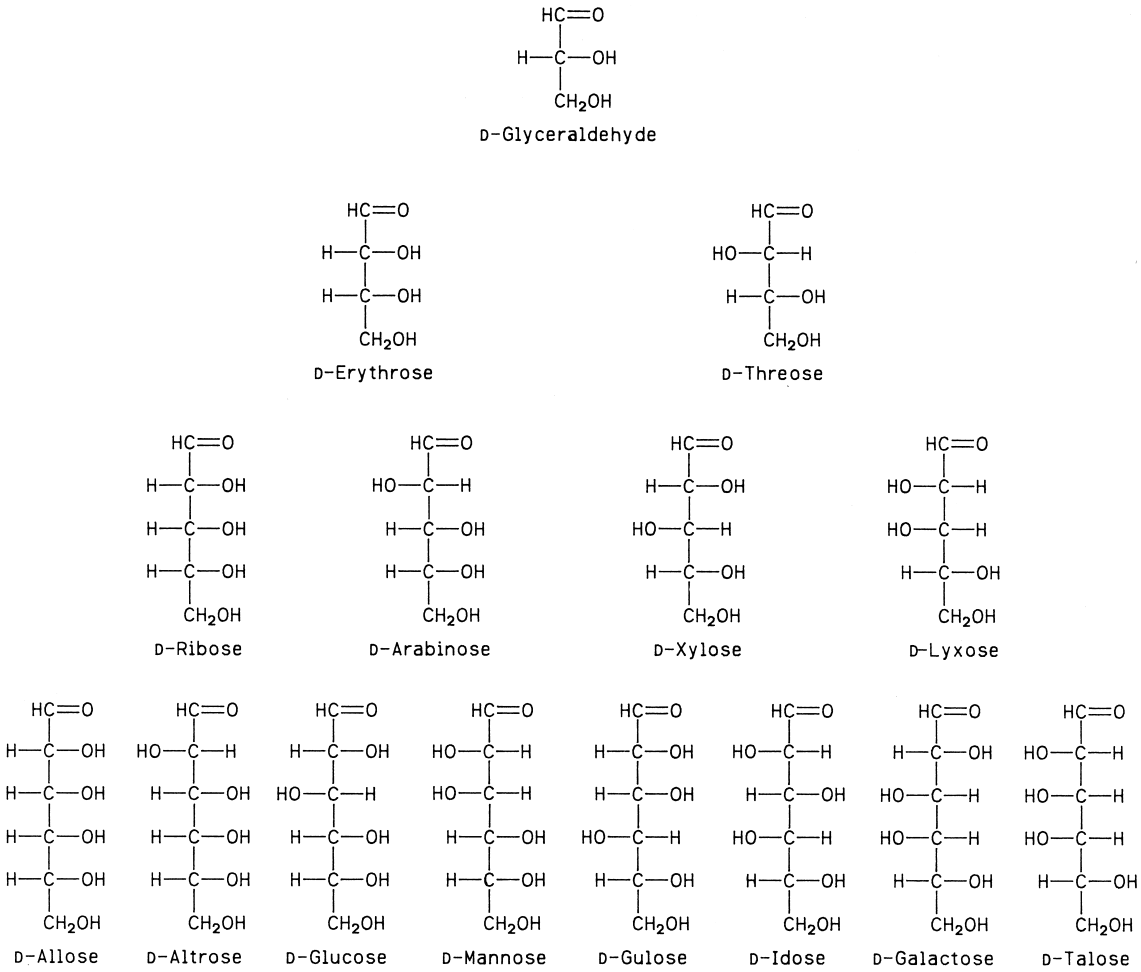


FIGURE 4 Aldoses of the D family.

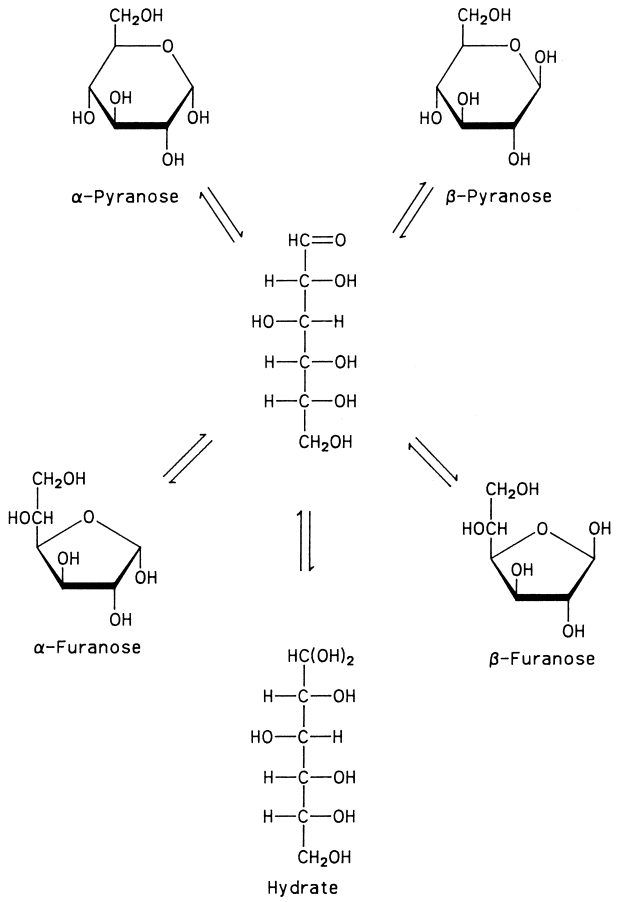
lacks an aldehydic proton at δ 10, also characteristic of such a group.

Two forms of D-glucose were isolated and designated, α and β (a third one was also isolated, but it proved to be an equilibrium mixture). Each of the two forms has, in solution, a characteristic optical rotation that changes with time until it reaches a constant value (that of the equilibrium mixture). The change in optical rotation with time is called *mutarotation*, and it is indicative of molecular rearrangements occurring in solution:

$$\begin{array}{ccccc} \alpha\text{-D-Glucose} & \rightleftharpoons & \text{mixture} & \rightleftharpoons & \beta\text{-D-glucose} \\ [\alpha]_D + 112.2^\circ & & +52.7^\circ & & +18.7^\circ \end{array}$$

It is now known that D-glucose and the other aldoses that have the necessary number of carbon atoms exist mainly in the form of five- or six-membered cyclic hemiacetals and rarely in seven-membered ones. These forms are the outcome of an intramolecular nucleophilic attack by the hydroxyl oxygen atom attached to C-4 or C-5 on the carbonyl group. The five-membered ring produced in the first case is related to tetrahydrofuran and is therefore designated *furanose*. The six-membered ring produced in the second case is related to tetrahydropyran and is called *pyranose*. Because cyclization converts an achiral aldehyde carbon atom to a chiral hemiacetal carbon atom, two isomers are produced, which have been designated α and β. Accordingly, a solution of an aldose at equilibrium contains a mixture of at least two (α and β) furanoses, two (α and β) pyranoses, as well as traces of the acyclic form and its hydrate (see Scheme 1).

Although usually there is a preponderance of the two pyranose forms, the composition of the equilibrium mixture and the contribution of each of the various forms differ from one sugar to another, depending on the instability factors dictated by the configuration. For example, because the interaction between cis OH groups on adjacent carbon atoms renders this arrangement considerably less desirable in a furanose ring than a pyranose ring, it was found that the all-*trans*-β-D-galactofuranose contributes more to the equilibrium (3%) than β-D-glucofuranose (0.1%).



SCHEME 1 Cyclic and acyclic forms of D-glucose existing in solution.

The anomeric configuration refers to the chirality at C-1 of a cyclic aldose or C-2 of a cyclic 2-ketose. These centers are achiral in the acyclic forms, and their chirality is the result of their conversion to cyclic hemiacetals. Two *anomeric* furanoses and two anomeric pyranoses can be produced from an acyclic monosaccharide, provided that the latter possesses the requisite number of atoms (four carbon atoms for a furanose ring and five for a pyranose ring). The configuration of the anomeric center, designated by the Greek letters α and β, is related to the last chiral center in the following way. In the D series, if the OH on the anomeric carbon is to the right in a Fischer projection or down in a Haworth formula (see below), the isomer is α, and if the OH is to the left or up, it is β. Conversely, in the L series, the α anomer has the C-1 OH to the left in a Fischer projection or up in a Haworth formula, and the β anomer has this OH to the right or down (see Fig. 5).

The common method of depicting monosaccharides in their cyclic forms, without assigning a specific conformation to the ring, is to use the Haworth formula. A pentagon having 108° angles or a hexagon with 120° angles is represented as seen by an observer situated at an angle of ~60° above the plane of ring, and as a precaution against optical illusions, regarding the side closer to the viewer, the bonds nearest to the observer are thickened, to give a sense of perspective (see Fig. 5).

Although rings may be turned around their centers, it is customary to orient them in such a manner that C-1 is to the right, and the ring oxygen is farthest from the viewer.

3. Conformation

Pyranose rings can exist in a number of inter-convertible conformers, of which the chair forms are the most stable. The number of recognized forms of the pyranose ring are two *chair* (C), six *boat* (B), four *half-chair* (H), six *skew* (S), and six *sofa* forms. (See Fig. 6 for the two chair and two of the boat forms.) To designate each of these forms, the number of the ring atom(s) lying above the plane of the pyranose ring is superscripted before the letter designating the form (C, B, H, S, etc.), and the number of the ring atom(s) lying below the plane is subscripted after the letter, e.g., $^{4}C_{1}$ (see Fig. 7).

It is possible to determine the conformation of a saccharide or glycoside either experimentally or by determining on purely theoretical grounds which conformer of a given ring (pyranose or furanose) will be the most stable.

If the conformation in the solid state is desired, X-ray crystallography or neutron diffraction is prescribed. It will provide the exact location of each atom in the crystal lattice and, by so doing, show the anomeric configuration and the conformation of the ring. X-ray crystallography and neutron diffraction afford unambiguous diagrams such as Fig. 8, which shows the conformation of methyl α-D-glucopyranoside as deduced from neutron diffraction analysis using a computer.

If, on the other hand, it is desired to know the conformation of a saccharide in a given solvent, NMR spectroscopy is used. Use is made of the fact that the anomeric proton of cyclic sugars and glycosides is the only one attached to two oxygen atoms, which deshield it. Accordingly, an ^{1}H NMR spectrum of such compounds will reveal at low field a well-defined doublet (coupled by the H-2 proton), whose coupling constant $J_{1.2}$ will give the dihedral angle between protons 1 and 2. Then by using proton decoupling techniques, it is possible to identify H-2 and determine the dihedral angle between it and H-3, and so on. In this way it is possible to determine the relative orientation of the protons on successive pairs of carbon atoms, and by summing up the data, one can obtain the conformation of the whole molecule, as well as the anomeric configuration. Figure 9 shows the ^{1}H-NMR spectrum of α-D-glucopyranose.

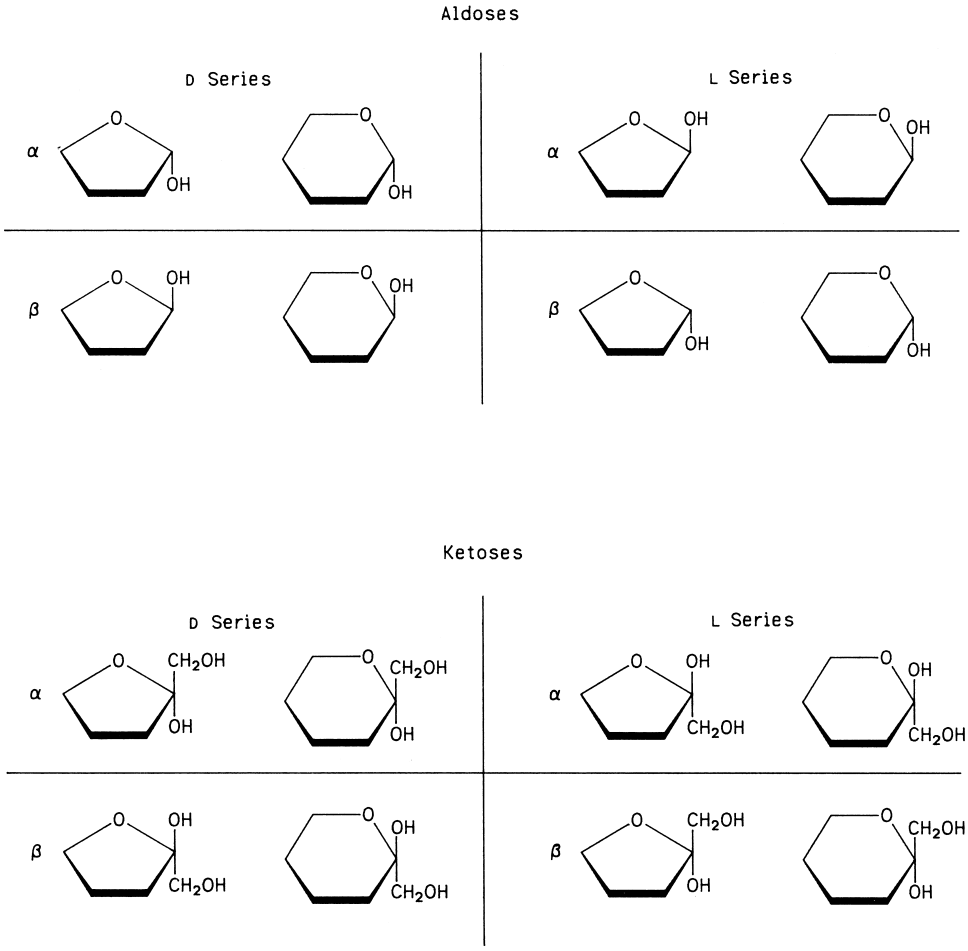


FIGURE 5 α and β configuration of aldoses and ketoses.

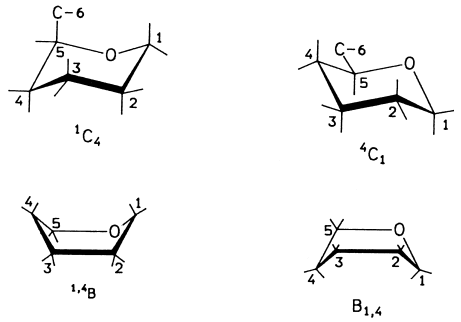


FIGURE 6 Two chair and two boat conformations.

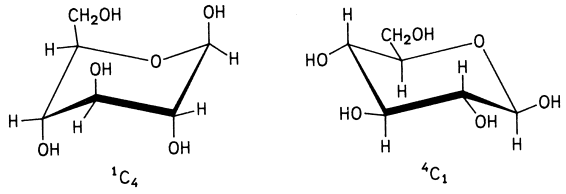


FIGURE 7 The two chair conformations of β-D-glucose.

Conformation analysis of the possible pyranose forms shows that the two chair conformations are so much more stable than the others that the analysis should be restricted to the choice between the $^4C_1$ and the $^1C_4$ conformers. Models of the two forms are checked for the presence of any of the instability factors listed in Table V and appropriate numerical value(s) assigned to each. These are then summed, and the form having the smallest numerical total is the more stable conformer.

The most important instability factor in the pyranose ring (with a value of 2.5 units) is a situation called delta 2. This is when the OH on C-2 bisects the angle between the ring oxygen atom and the oxygen atom of the anomeric hydroxyl group, forming an isosceles triangle having one oxygen atom at each corner. An instability factor of 2.0 units is given to an axial CH_2OH. An axial hydroxyl group on the ring has an instability factor of 1.0 unit. The 1,3 interaction between an axial CH_2OH and an axial OH group on the same side of the ring produces an unfavorable situation, of an additional 2.5 units.

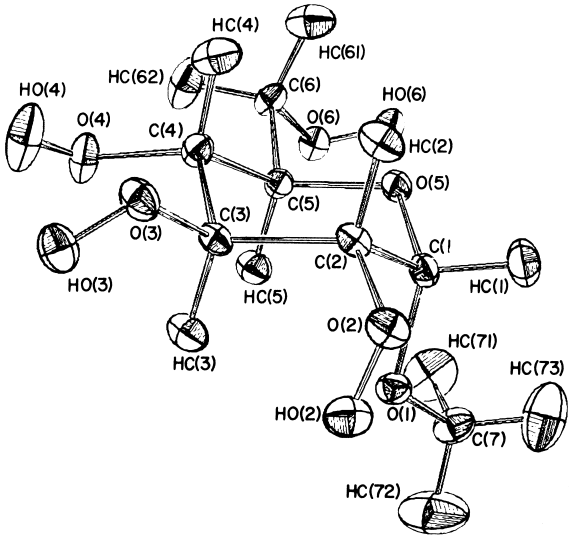


FIGURE 8 Neutron diffraction of methyl α-D-glucopyranoside.

TABLE V Free Energy of Unfavorable Interactions

Type of interaction	Free energy (kJ)
Gauche-1,2	
O-1-*e*-O-*e*	2.30
O-1-*e*-O-*a*	4.19
O-*e*-O-*e* or -*a*	1.47
C-*e*-O-*e* or -*a*	1.88
Axial–axial-1.3	
O–O	6.28
O–C	10.47
O–H	1.88
C–H	3.77
Anomeric effect	1.26

The anomeric effect destabilizes the anomer having an equatorial hydroxyl group on C-1 (e.g., β-D-glucopyranose) due to the dipole–dipole interactions between its unbonded electrons and the unbonded electrons of the ring oxygen atom. The anomeric effect depends on the polarity of the solvent, and this is why one does not give it a high numerical value when computing the instability factors of free sugars dissolved in water. It does, however, play an important role during glycoside formation, favoring the conformer having an axial OR group on C-1, such as an α-D-glucopyranoside. Empirical calculations of the conformational free energy by summation of the unfavorable interactions have been used successfully in conformational analysis to determine the more stable conformer of a pyranose in a chair form. Two types of interaction are used: those between two gauche-1,2 substituents [either equatorial–equatorial (*ee*) or axial–equatorial (*ae*)], and those between two syn-diaxial-1,3 groups.

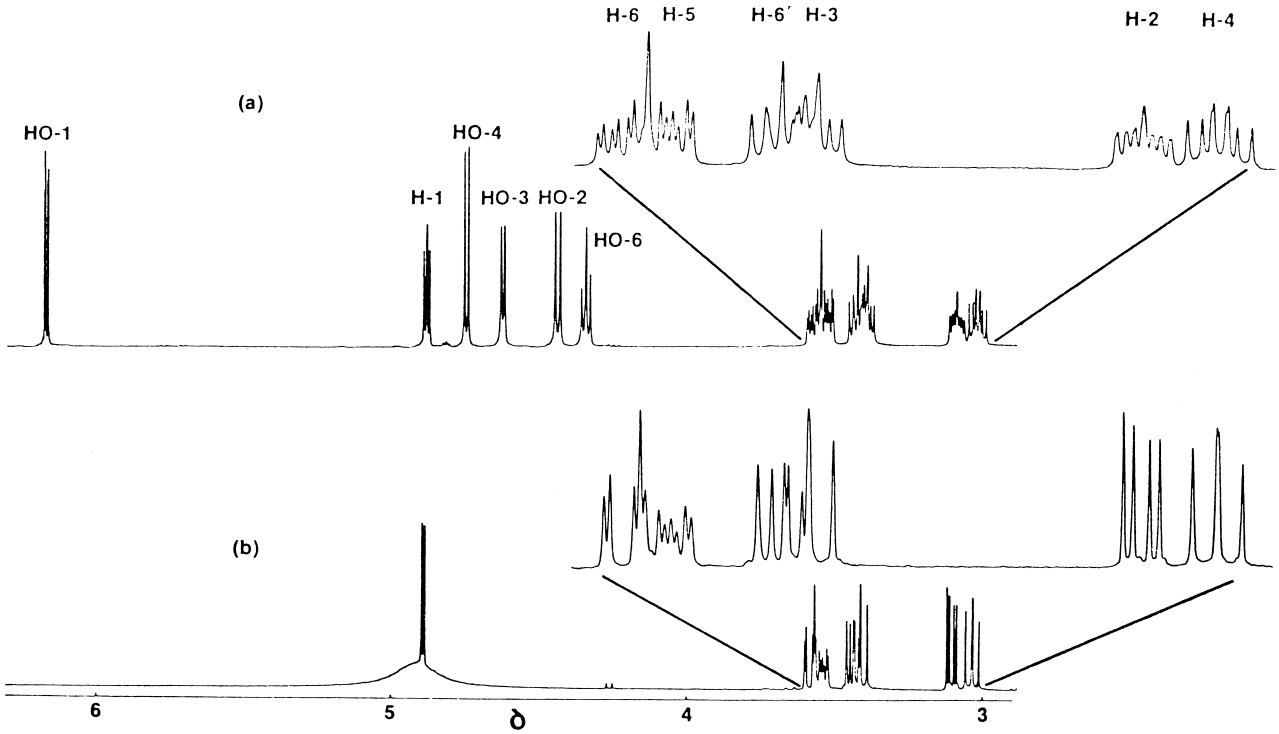


FIGURE 9 ^{1}H NMR spectra of a solution of α-D-glucose in DMSO-d_6 at 400 MHz: (a) OH protons coupled and (b) OH protons decoupled.

TABLE VI Stable conformation of D-Aldopyranoses in Aqueous Solutions

Aldose	Conformation
Aldohexoses	
α-D-Allose	$^4C_1$
β-D-Allose	$^4C_1$
α-D-Altrose	$^4C_1$, $^4C_4$
β-D-Altrose	$^4C_1$
α-D-Galactose	$^4C_1$
β-D-Galactose	$^4C_1$
α-D-Glucose	$^4C_1$
β-D-Glucose	$^4C_1$
α-D-Gulose	$^4C_1$
β-D-Gulose	$^4C_1$
α-D-Idose	$^4C_1$, $^1C_4$
β-D-Idose	$^4C_1$
α-D-Mannose	$^4C_1$
β-D-Mannose	$^4C_1$
α-D-Talose	$^4C_1$
β-D-Talose	$^4C_1$
Aldopentoses	
α-D-Arabinose	$^1C_4$
β-D-Arabinose	$^4C_1$, $^1C_4$
α-D-Lyxose	$^4C_1$, $^1C_4$
β-D-Lyxose	$^4C_1$
α-D-Ribose	$^4C_1$, $^1C_4$
β-D-Ribose	$^4C_1$, $^1C_4$
α-D-Xylose	$^4C_1$
β-D-Xylose	$^4C_1$

Using these values, it can be calculated that the $^4C_1$ form of β-D-glucopyranose has a conformational energy of 8.38, considerably lower than that of the $^1C_4$ form with a value of 33.5, leaving no doubt that the first is the more stable conformer. Table VI shows the conformation of D-aldopyranoses in aqueous solutions.

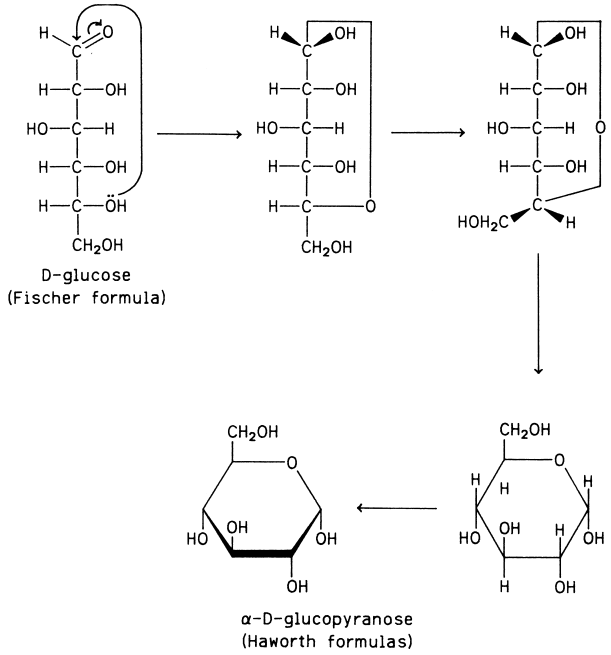
The principal conformers of the furanose ring are the *envelope* (E) and the *twist* (T) forms, of which the latter are the most stable. To designate a particular conformation, the method is the same as that used for pyranoses; namely, the letter used to designate the form (E or T) is superseded by the number of the ring atom situated above the plane of the ring and is followed by the number of the ring atom below the plane of the ring.

For furanose rings, the most stable conformers are the envelope and twist forms; these can exist in 10 arrangements each. Due to the low energy barriers between the twist and envelope conformers, a sugar in the furanose form is believed to undergo rapid interconversions between them. The slightly more favored twist conformer would rapidly pass through an envelope conformation (which is less favored because it possesses two eclipsed carbon atoms) to go to the next twist form. Because interaction between two eclipsed carbon atoms is greater than between a carbon atom and an oxygen atom, the ring oxygen atom tends to occupy a position along the plane of the ring and to leave the puckering to carbon atoms.

B. Reactions of Monosaccharides

In reactions involving monosaccharides, it is important to remember that the functional groups found in the various cyclic and acyclic forms will be present side by side in the reaction mixture. Thus, in a reaction involving an aldohexose, a carbonyl group and two types of hydroxyl groups will be provided by the acyclic form. These are the primary hydroxyl group attached to the terminal position, and four (less reactive) secondary hydroxyl groups. In addition, each of the cyclic forms will contribute three types of hydroxyl groups: a hemiacetal hydroxyl function at C-1, a terminal primary hydroxyl group, and three (less reactive) secondary hydroxyl groups.

It is also important to remember that the course of a reaction does not depend totally on the relative amount of a given species. Thus, free saccharides readily undergo nucleophilic additions, characteristic of carbonyl groups, even though carbonyl groups are present only in their acyclic form, which contributes little to the equilibrium



SCHEME 2 Representation of the anomeric configuration D-glucose in a Fischer (top right) and a Haworth (bottom left) formula.

mixtures. For example, solutions of hexoses contain less than 1% of the acyclic form, and yet they readily afford carbonyl group derivatives in high yields. This is because the addition of a nucleophile to the carbonyl group of an acyclic form will immediately shift the equilibrium in favor of this form, allowing more of it to be formed and to react with the nucleophile.

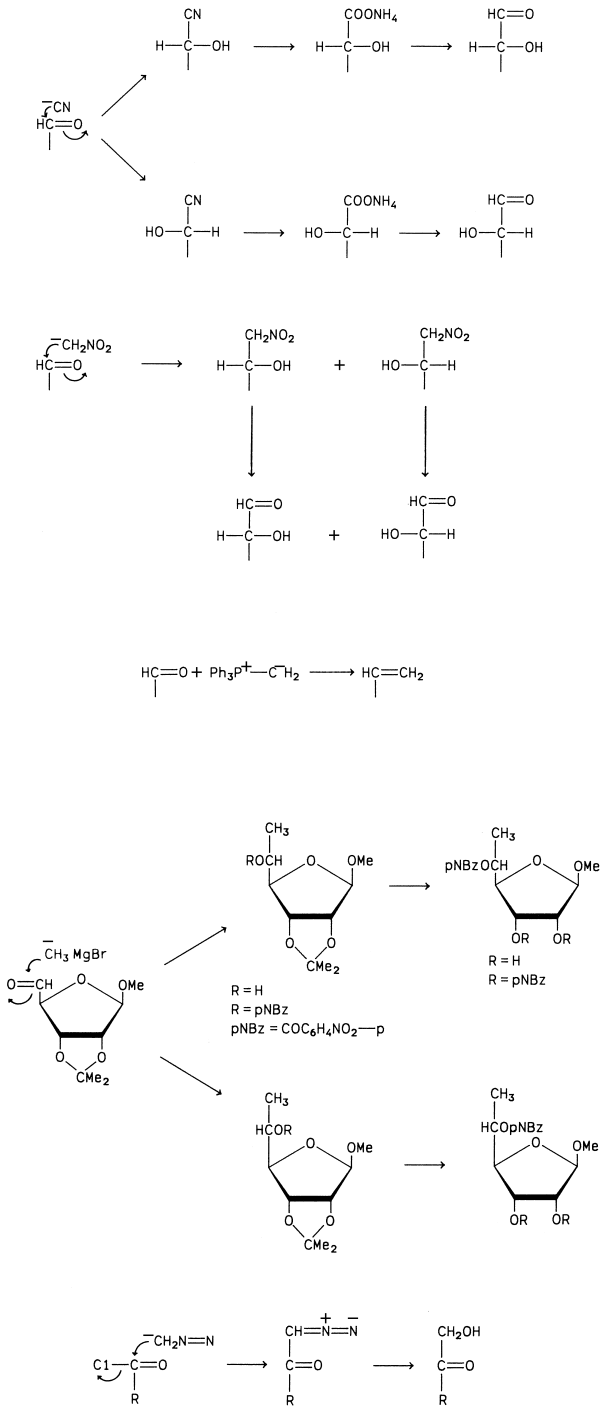
Concerning the hydroxyl groups, it should be recognized that the hydroxyl groups can play a dual role. Such groups can act as leaving groups when stronger nucleophiles attack the carbon atom to which they are attached. These nucleophilic substitution reactions may be of the S_N1 or S_N2 type. Alternatively, the oxygen of the hydroxyl group acts as a nucleophile, adding to carbonyl groups and carbonium ions or displacing good leaving groups to afford esters, acetals, ketals and ethers, and other species. The hydroxyl group of the hemiacetal function at C-1 of an aldose or C-2 of a ketose is the most reactive of all the hydroxyl groups found in a monosaccharide. The next most reactive hydroxyl group is the primary hydroxyl group at the terminal position. This is followed in reactivity by the secondary hydroxyl groups. The oxidation of carbonyl and hydroxyl groups has important applications in industrial processes.

1. Reactions of the Carbonyl Group

The reactions of the carbonyl group of a monosaccharide include the nucleophilic addition of a carbon, nitrogen, oxygen, or sulfur atom. It should be noted that, although these additions afford acyclic products, the latter may cyclize, so that the product may ultimately be acyclic or cyclic. On the other hand, intramolecular nucleophilic addition, by a hydroxyl group attached to the sugar chain on the carbonyl carbon atom, can afford only cyclic products. Note also that, in both types of additions, the reaction may occur with or without subsequent loss of water.

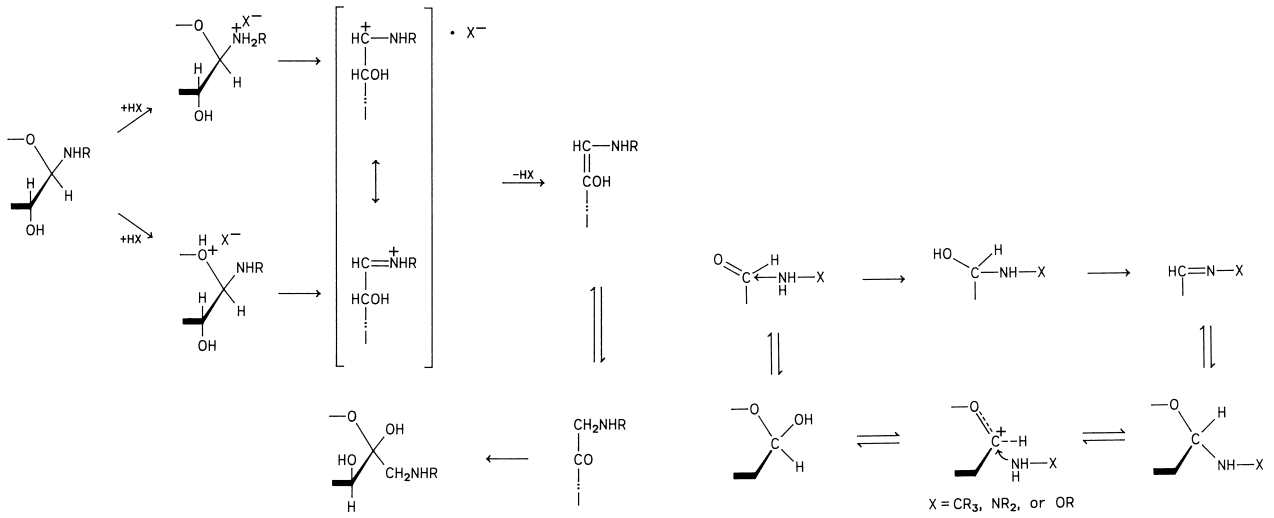
The addition of carbon nucleophiles to the carbonyl group of aldoses has been widely used to extend the carbon chains of saccharides, that is, to ascend the series. Similar additions to the keto group of glyculoses have been used to prepare branched sugars. Of particular value in forming C–C bonds are the nucleophiles ^-CN, $^-CH_2NO_2$, and $^-CH_2N_2$, as well as the ylides and organometallic nucleophiles involved in the Wittig and Grignard reactions (Scheme 3).

The nitrogen nucleophiles commonly used with monosaccharides contain a primary amino group attached to a carbon, a nitrogen, or an oxygen atom. The products obtained from the first type of nucleophile are rarely of the acyclic, Schiff base type, because they readily cyclize. The other two types of nucleophile afford carbonyl-group derivatives that usually exist in acyclic forms. The major



SCHEME 3 Addition of carbon nucleophiles to carbonyl groups.

pathway for the reaction between a nitrogen nucleophile and a free sugar is via the acyclic form of the monosaccharide, which usually affords an acyclic addition product. The latter may subsequently lose water to yield an acyclic condensation product, or it may cyclize. The same reaction product may be formed by nucleophilic substitution



SCHEME 4 Amadori rearrangement of aldimines via eneaminols.

on the hemiacetal function of a cyclic sugar. Nucleophlic substitution is much slower than nucleophilic addition, but in this case its contribution is enhanced by the large proportions of cyclic forms of the sugar present in the equilibrium mixture.

Aldoses and ketoses react with ammonia and amines to give 1-deoxy-1-imino- and 2-deoxy-2-imino derivatives, respectively, both of which exist mainly in cyclic forms, referred to as glycosylamines. During the reaction of aldoses with amines, the 1-amino-1-deoxyuloses, that is, glycosylamines that are first formed, often rearrange to give 1-amino-1-deoxyketoses, called *Amadori compounds*. The reaction leading to their formation is also named after its discoerer and referred to as the *Amadori rearrangement*. Unlike glycosylamines, which exist preponderantly in cyclic forms, Amadori compounds may be cyclic or acyclic. The mechanism of the reaction has been extensively studied by Weygand, who proposed the mechanism shown in Scheme 4, which involves a sigmatropic rearrangement of an aldimine to a 1,2-enaminol, which later ketonizes.

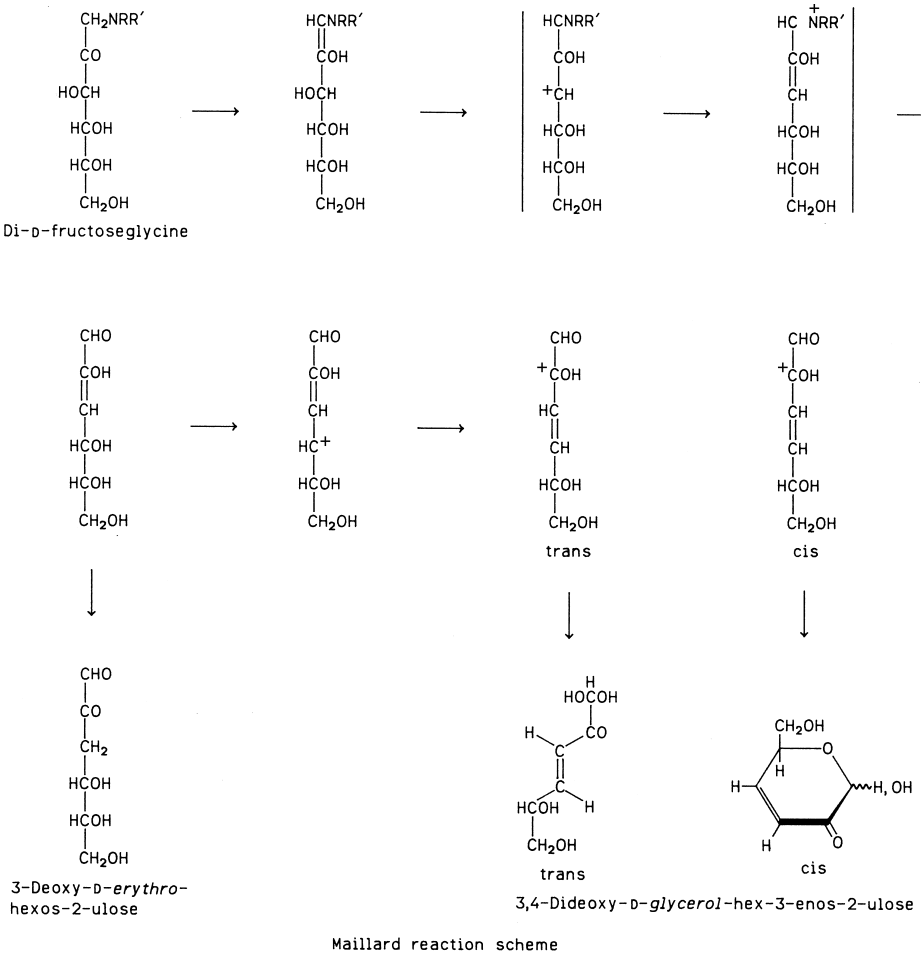
Sigmatropic rearrangements leading to the formation of enaminols are by no means restricted to imines, as they occur during the conversion of hydrazones into osazones, and are involved in the cyclization of the latter compounds.

The analogous reaction between monosaccharides and amino acids or peptides is of great importance to the food industry. This is because its ultimate outcome is the formation of the dark polymeric products known as *malanoidins*, which give baked goods their characteristic color. The initial stages of this complicated reaction, collectively known as the *Maillard reaction* (Scheme 5), are well documented. They involve the formation of *N*-glycosylamino acids, which undergo Amadori rearrangement to ketose amino acid derivatives and then dimerize to diketose amino acids. The latter then undergo a series of double-bond migrations, via 1,2-enolizations and 2,3-enolizations, to afford mono- and dideoxyhexosuloses. The glycosuloses are then attacked by amino acids and, after undergoing polymerizations and decarboxylations, yield polymeric melanoidins.

The hydrazine derivatives of sugars are more reactive than the sugars from which they are prepared. This is because, when a hydrazone residue is introduced into an aldose molecule, the number of nucleophilic groups that are capable of entering a reaction is increased by unity (the second nitrogen atom of the hydrazone acts as a strong nucleophile), whereas the number of groups capable of undergoing nucleophilic attack remains constant because nucleophiles will add to the C═N group at position 1 in the same way that they do to the C═O group of the parent sugar. In the case of osazones and other bis(hydrazones), the reactivity is further enhanced because a second C═N group is introduced (in place of a less reactive H─C─OH group). As a result, the capacity of osazones and bis(hydrazones) to undergo addition and cyclization reactions is greater than that of monohydrazones, which in turn is greater than that of the free sugars that yielded them.

Saccharide hydrazones have been prepared from unsubstituted, monosubstituted, and N,N-disubstituted hydrazines. The substituents attached to the hydrazine include alkyl, aryl, or heteroaryl groups, as well as acyl, aroyl, thioacyl, thioaroyl, and sulfonyl groups.

Sugar hydrazones can exist as equilibrium mixtures of various tautomeric forms. This is apparent from the complex mutarotation they exhibit in solution. The most

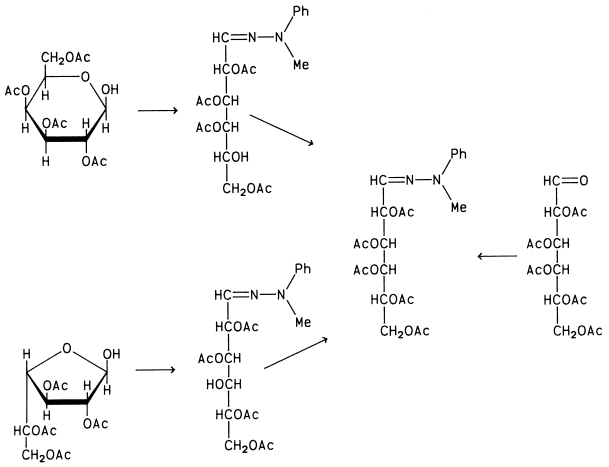


SCHEME 5 Millard reaction scheme.

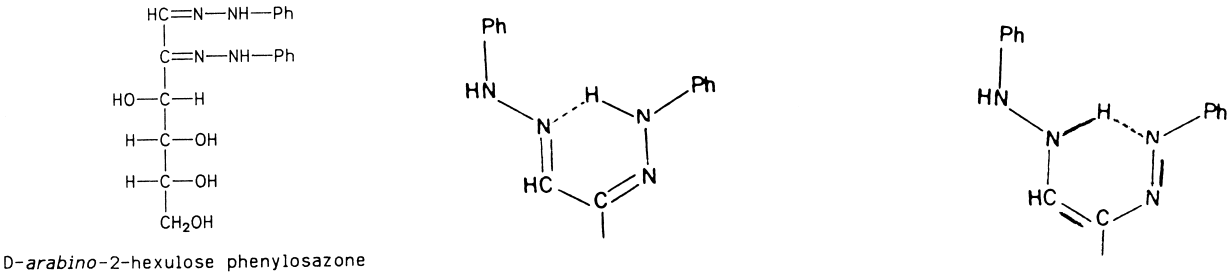
important of these structures are the acyclic Schiff bases and the two pairs of anomeric cyclic forms (the two five- and two six-membered rings). However, when crystallized, one form of the hydrazone is usually isolated. For example, the crystalline form of D-galactose phenylhydrazone was shown by the following sequence of reactions to be the acyclic form (see Scheme 6).

Osazones and bis(hydrazones) are hydrazine derivatives of aldosuloses and diuloses, respectively. If the two hydrazone residues are attached to C-1 and C-2 of a saccharide, the derivative is referred to as an osazone and is named by suffixing osazone to the name of the ketose that possesses the same carbon chain and the same configuration (irrespective of the sugar used in its preparation); for example, D-*arabino*-2-hexulose phenylosazone is the name given to the osazone obtained from D-glucose, D-mannose, and D-fructose. The stability of osazones is attributed to their chelated rings (see Scheme 7).

Reducing sugars react with hydroxylamine to give oximes, which have been prepared for use in the Wohl



SCHEME 6 Formation of the same galactose hydrazone pentaacetate from a pyranose, a furanose, and an acyclic sugar acetate.



D-*arabino*-2-hexulose phenylosazone

SCHEME 7 Chelated ring structures of D-*arabino*-hexulose phenylosazone, represented by a Fischer formula on the left.

degradation for descending the series. This reaction makes use of the fact that, on refluxing with acetic anhydride, the oxime readily loses water (or acetic acid, if it is first N-acetylated) and is converted to the nitrile. Then using the fact that the addition of HCN to aldehydes is reversible, the nitrile formed is hydrolyzed to an aldose having one carbon atom less than the starting oxime (see Scheme 8). If carried out with ammonia, the hydrolysis affords an undesired intermediate bis(acetamido) derivative, which must by hydrolyzed with acid before the desired aldose can be liberated.

The addition of l mol of RO^- and RS^- nucleophiles to the carbonyl group of a saccharide affords the hemiacetal or thiohemiacetal, which in solution remains in equilibrium with the free saccharide. On the other hand, the addition of 2 mol of the nucleophile affords stable acetals or dithioacetals (see Fig. 11).

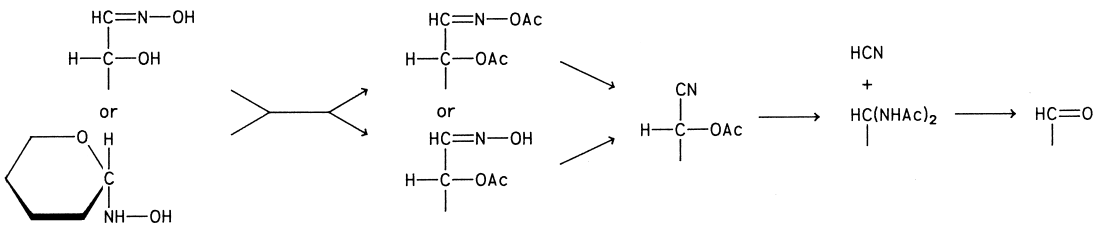
If one of the OH groups of the acyclic form of the sugar is the nucleophile, the product is one of the four possible cyclic structures of a sugar (α- and β- furanose and α- and β-pyranose). If, instead, the nucleophile originates from without the sugar molecule, hemiacetals or thiohemiacetals are obtained, depending on the nucleophile. Because these compounds are constantly in equilibrium with the acyclic form, they are not isolated and have little importance in preparative chemistry. Three types of acetal are possible, depending on the source of the nucleophiles(s) involved in the reaction: (1) If both nucleophile molecules do not arise from the sugar molecule, the product is an acyclic acetal or dithioacetal; (2) if one nucleophile forms part of the sugar molecule and one is not a part of it, there is obtained a cyclic acetal or thioacetal, which is referred to as a glycoside or thioglycoside; (3) if both nucleophile molecules are part of the same sugar molecule, the product is a bicyclic acetal. Acetals of types 2 and 3 are, in reality, derivatives of the cyclic forms of sugars, and, as such, they do not belong among the reactions of the carbonyl group.

In general, the best route for preparing the acyclic forms of sugars and acyclic sugar derivatives is that from the dithioacetals; another way is to prepare an acyclic hydrazone of the sugar, acetylate it, and then remove the hydrazone residue from the acetate with benzaldehyde. Because of the importance of dithioacetals in synthesis, their preparation, as well as their conversion to acetals and other useful derivatives, will be examined. Dithioacetals are readily obtained in acidic media by the treatment of monosaccharides with the desired thio alcohol, usually ethanethiol, in the presence of concentrated HCl (see Scheme 9).

The demercaptalation (removal of the thioacetal residues) is usually conducted in the presence of a mixture of yellow mercuric oxide and cadmium carbonate or mercury chloride. If hydrolysis is needed in order to prepare the aldehydo sugar, water is added to the mixture, care being taken that the reaction is performed on the peracetylated dithioacetal (to prevent formation of the cyclic forms).

2. Nucleophilicity Quotient of Monosaccharides

The nucleophilicity quotient (NQ) is a measure of the susceptibility of polyfunctional molecules to cyclization by intramolecular, nucleophilic addition. Sugars and sugar



SCHEME 8 The Wohl degradation is used to reduce the number of C atoms in sugar.

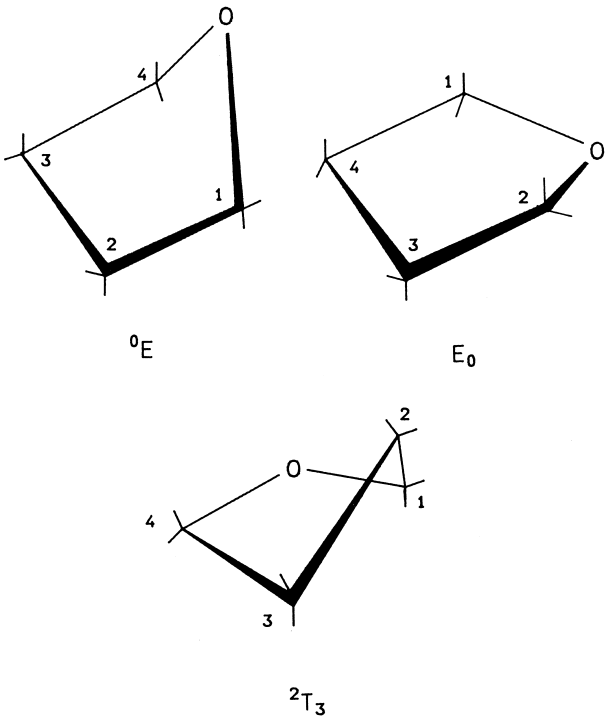


FIGURE 10 Envelope and twist conformations of furanoses.

derivatives possess more nucleophilic species than nucleophile acceptors (e.g., there are five nucleophilic oxygen atoms vs. one carbonyl group in the acyclic form of an aldohexose); accordingly, the number of nucleophile acceptors exerts a greater, and sometimes controlling, influence on the capacity of the molecule to cyclize, whereas the nucleophilicity of the attacking groups plays a lesser role. For this reason, integers are used to designate the number of nucleophile acceptors (such groups as C═O, C═NR, C═C─C═O, and C═C─C═NR, which can undergo addition reactions), and fractions are used to designate the nucleophilicity (n) of attacking species (e.g., OH or NH groups) that are suitably situated to react with a nucleophile acceptor. The value of n, which is a measure of the affinity of the nucleophile to a particular acceptor, can be obtained from tables or can be determined experimentally from kinetic measurements using a Hammett-like equation, namely, $\log(k/k_0) = ns$, where s is the sensitivity of the nucleophile acceptor to the attacking nucleophile. Free aldoses have an NQ value of 1.4, which signifies that (1)

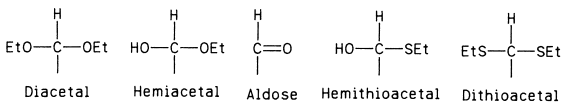
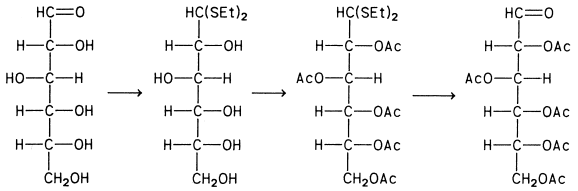


FIGURE 11 Acetals hemiacetals, and their sulfer (thio) derivatives.



SCHEME 9 Formation of acyclic aldoses from diothioacetals.

these molecules possess one nucleophile acceptor (C-1 of the acyclic form; the carbon atoms in positions 2, 3, 4, 5, and 6 are not counted, because their susceptibility to nucleophiles by S_N reactions is much smaller than that of C-1, which undergoes carbonyl addition reactions) and (2) the nucleophilicity of the attacking oxygen atom (as determined from tables) is 4. Saccharide hydrazones and oximes have similar NQ values, even though they possess additional nucleophiles (NH or OH groups) whose nucleophilicity is enhanced by adjacent nitrogen atoms. This is because these additional nucleophiles are not suitably situated to attack the nucleophile acceptor (the C═N group). Saccharide phenylosazones, on the other hand, are much more reactive, and this is reflected by the high NQ value of 2.6. They possess two nucleophile acceptors, namely, C-1 and C-2, which form part of C═NR systems, and contain a nitrogen atom, which is suitably situated to attack a nucleophile acceptor (the C═C─C═N group present in one of the tautomeric forms in equilibrium). The nucleophilicity of this atom is enhanced by the α effect of the adjacent nitrogen atom, giving it an n value of 6. Finally, if it is desired to obtain compounds that are even more reactive, another nucleophile acceptor could be added to the osazone molecule, for example, a carbonyl group in the saccharide moiety or in the hydrazone residue, which would bring their NQ values to 3.6. Examples of such highly reactive molecules are L-ascorbic acid bis(phenylhydrazones) and glycosulose bis(benzoylhydrazones).

3. Nucleophilic Substitution Reactions of the Anomeric Carbon Atom

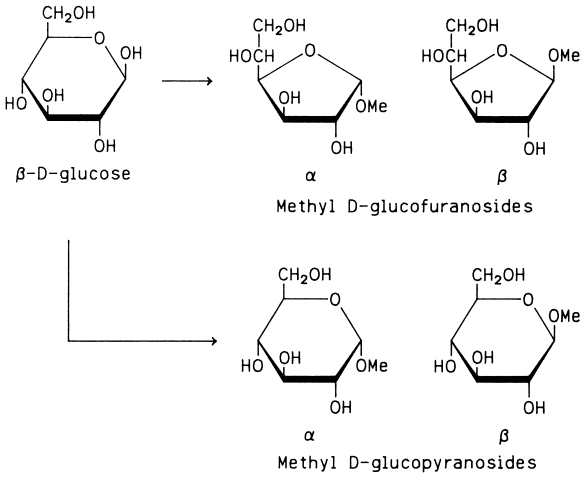
The anomeric carbon atom is the most reactive carbon atom in furanoses, pyranoses, and their derivatives. Although it is less susceptible to nucleophilic attack than the carbonyl group of acyclic sugars (because it undergoes nucleophilic substitution instead of addition), it is significantly more reactive than the remaining hydroxyl-bearing carbon atoms. The affinity of the anomeric carbon atom to nucleophiles is attributed to its capacity to form resonance-stabilized carbonium ions that can undergo direct reactions via S_N1 mechanisms or synchronous reactions by S_N2 mechanisms. The rates of these reactions

depend on the nucleophilicity of the attacking group and the nature of the leaving group, that is, how good a leaving group it is.

The nucleophilic substitution reactions that will be considered in this section are (1) the displacement of the hemiacetal hydroxyl group (—OH) or one of its esters, for example, an anomeric acetoxyl group (—O—CO—R), by an OR group from an alcohol to afford glycosides, or by the X group of hydrogen halides to afford glycosyl halides; and (2) the displacement of the OR groups of glycosides by OH groups (hydrolysis), by OR groups (anomerization and transglycosidation), or by a halogen.

a. Displacement of OH groups by OR groups (glycosidation). The anomeric hydroxyl groups of cyclic sugars and their esters can be exchanged in acid media by the OR group of alcohols and phenols. The displacement of the hemiacetal hydroxyl group by an OR group is known as the Fischer method of glycoside formation, and the similar exchange of an anomeric ester group is referred to as the Helferich method.

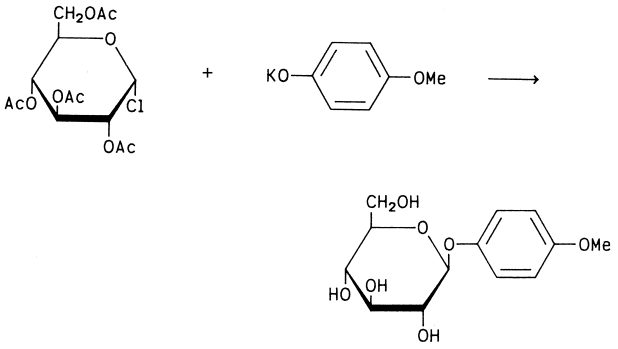
i. The Fischer glycosidation method. An example of the Fischer glycosidation method is the conversion of D-glucose to two methyl D-glucofuranosides and two methyl D-glucopyranosides by treatment with methanolic HCl. The first products isolated are two furanosides, methyl α-D-glucofuranoside and methyl β-D-glucofuranoside. The furanosides are kinetically favored because closure of five-membered rings is faster than that of six-membered rings. If the reaction time is prolonged or if reflux temperatures are used, the thermodynamically favored pyranosides will predominate, affording methyl α-D-glucopyranoside and methyl β-D-glucopyranoside in high yields (see Scheme 10). A study of the reaction



SCHEME 10 Formation of methyl α and β-D-glucofuranosides and methyl α and β-D-glucopyranosides.

products of the Fischer reaction revealed the following points. (1) At equilibrium, there is usually a preponderance of pyranose forms, since the six-membered rings are the thermodynamically favored forms of glycosides. (2) As a rule, aldoses having adjacent bulky groups (OH or OMe groups) in the cis orientation will afford a smaller proportion furanosides on equilibration. This is because these substituents are closer together (and repel one another more) in furanosides than in pyranosides. For the same reason, a trans relationship of adjacent bulky groups is more favorable in furanosides than in pyranosides. This is why furanosides having C-1 and C-2 trans (e.g., the α anomers of D-arabinose and D-lyxose and the β anomers of D-ribose and D-xylose) are more stable and exceed the concentration of their anomers. It was also found that the all-*trans*-methyl β-D-arabinofuranoside is present in high concentration in the equilibrium mixture. In the case of the methylated derivatives of D-arabinose (which have bulkier OMe groups), the all-trans orientation is so favorable in the five-membered ring that, at equilibrium, the concentration of the furanosides exceeds that of the pyranosides. (3) Because of the anomeric effect, the anomer having an axial methoxyl group always predominates among the two pyranosides (in furanosides no substituent exists in a truly axial position). (4) During the formation of methyl glycosides by the Fischer method, a small proportion of the acyclic acetal is always produced. This by-product is formed by nucleophilic addition to the carbonyl group of the acyclic sugar, as well as by nucleophilic substitution on the furanose form of the sugar, which can cause the ring to open.

ii. The Helferich glycosidation method. The Helferich method of glycoside formation involves allowing a peracetylated sugar (obtained by treating a free sugar with acetic anhydride in pyridine) to react with a phenol or an alcohol in the presence of an acid catalyst, a Lewis acid such as $ZnCl_2$ or, a protic acid as *p*-toluenesulfonic acid. Because peracetylated sugars usually exist in the pyranoid form, this method constitutes a convenient way in which to prepare glycopyranosides (see Scheme 11). However, if a peracetylated furanose is the starting compound, the product will, evidently, be a glycofuranoside. If mild conditions are used (e.g., a short heating time with *p*-toluenesulfonic acid), the conversion can be achieved with retention of configuration. However, if drastic conditions are used, anomerization occurs, and that anomer having a bulky group axial on C-1 will predominate. As in the Fischer method, the intermediate here is a resonance-stabilized carbonium ion. Because halogens are good leaving groups, the glycosyl halides are important intermediates in carbohydrate synthesis. The most important derivatives are the glycosyl chlorides and bromides; the iodides are usually too reactive and therefore not stable over any length of time; and the



SCHEME 11 Formation of a phenylglucoside from a glycosyl chloride.

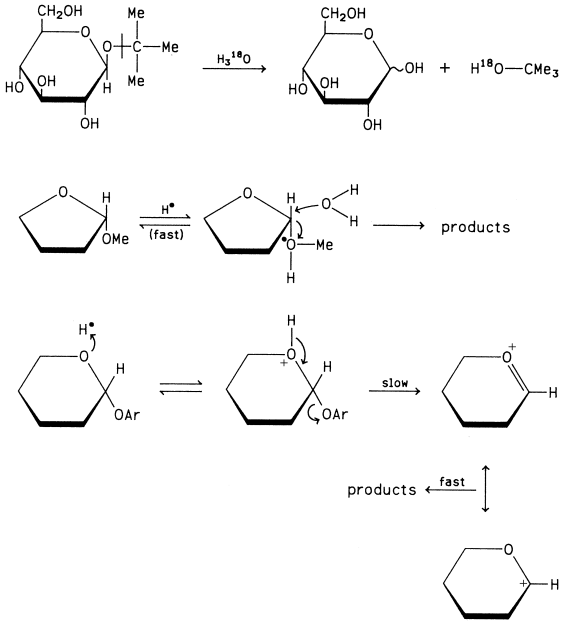
fluorides are too inert, requiring drastic conditions for reaction. The glycosyl halides can be prepared by bubbling HCl or HBr into a solution of a peracetylated or perbenzoylated monosaccharide. Alternatively, as will be seen later, they can be obtained in the same way from glycosides. Although many halides are isolated in crystalline form, they are often caused to react directly with a nucleophile after removal of the dissolved acid.

b. Nucleophilic displacement of OR groups. Nucleophilic substitution reactions on the anomeric carbon atom include displacement of the anomeric OH groups, acetoxyl groups and halogens but may involve displacement of the OR group of glycosides by a nucleophile. The nucleophiles discussed in this section are the OH and OR groups, as well as halides. Acid catalysts are always needed in the anomeric displacement reactions in order to produce the reactive species, a resonance-stabilized carbonium ion. The latter is usually formed when the oxygen atom of the OR group attached to the anomeric carbon atom becomes protonated and the aglycon is eliminated as ROH. If the glycoside reacts with the same nucleophile found in its aglycon (e.g., if a methyl glycoside is treated with methanol in the presence of an acid catalyst), anomerization will result, that is, the two anomers of the same glycoside will be produced, and the anomer having the OR group axially attached will predominate (the anomeric effect). If, on the other hand, the glycoside reacts with a different nucleophile, for example, if a methyl glycoside is treated with water (in a hydrolysis) or with another alcohol (in a transglycosidation), the OR group will be replaced by an OH group in the first case and by an OR′ group in the second. In such reactions, if the OH group attached to C-4 of glycopyranosides and C-5 glycofuranosides is not blocked, the foregoing displacement reactions will afford mixtures of α- and β-furanoses (or furanosides) and α- and β-pyranoses (or pyranosides), irrespective of whether the starting glycoside is a furanoside or a pyranoside.

Another important nucleophile that can react with glycosides is X^-, to produce glycosyl halides, which are synthetically valuable intermediates. In general, glycosides are hydrolyzed by acids and are stable toward bases. Exceptions to this rule are glycosides having, as the aglycon, a group derived from a phenol, an enol, or an alcohol having an electronegative group in the β position, as these are labile toward bases.

Aldofuranosides are hydrolyzed much faster (50–200 times the rate) than the aldopyranosides, and, among the furanosides, the less stable isomers (those having adjacent bulky groups in the cis orientation) are hydrolyzed the fastest. Similarly, methyl β-D-glycopyranosides are hydrolyzed faster than their (more stable) α-D anomers, because of the anomeric effect. (The situation is reversed when bulky aglycons are used, because of the considerable axial–axial interactions.) Ketofuranosides and ketopyranosides are also hydrolyzed faster than the corresponding aldosides. Finally, aldopyranoses and aldofuranoses having a 2- or 3-deoxy or a 2,3-dideoxy functionality are hydrolyzed considerably faster than their hydroxylated counterparts; this is due to a decrease in steric interaction during the conversion of the chair conformer of the glycopyranoside to the half-chair conformer of the carbonium ion and during the bimolecular displacement of the alcohol by water in the glycofuranosides.

The accepted mechanism for the acid hydrolysis of glycosides starts, as usual, with protonation. Although protonation of either the glycosidic oxygen atom (that of the —OR group attached to the anomeric carbon atom) or the ring oxygen atom is possible, there is evidence that it is usually the former (the glycosidic oxygen atom) that is protonated. The next step in the reaction, namely, the formation of a carbonium ion by shifting of the electrons of the C—O—R bonds, can be achieved in two ways: either by breaking the C—O bond and shifting the positive charge to the anomeric carbon atom or by breaking the O—R bond and shifting the charge to the R group of the aglycon. The first route is favored, because the saccharide carbonium ion is stabilized by resonance with the form having the charge on the ring oxygen atom, and most hydrolyses follow this route. In rare cases, for example, during the acid hydrolysis of *tert*-butyl glycosides, the bond between the oxygen atom and the R group is preferentially broken, because of the remarkable stability of the *tert*-butyl carbocation. The foregoing mechanisms have been confirmed by conducting hydrolyses of glycosides in ^{18}O-labeled water and locating the labeled oxygen atom in the products (the free sugar or the alcohol). With most aglycons the label was found to be attached to the free sugar liberated, whereas with *tert*-butyl glycosides the label was found to be on the alcohol formed (see Scheme 12).

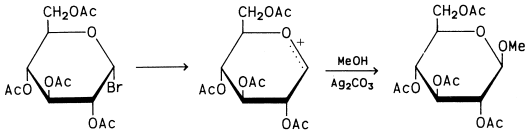


SCHEME 12 Mechanism of hydrolysis of glycosides.

Anomerization is the acid-catalyzed isomerization of a group (OR) attached to the anomeric carbon atom of a cyclic saccharide derivative. The reaction is initiated by protonation of the OR group of the glycoside and is followed by elimination of the alcohol group (ROH) to afford a carbocation. This then undergoes nucleophilic attack by an OR group (identical to the OR originally present in the glycoside). Of the two anomers possible, that having the bulkier axial group predominates (the anomeric effect).

An important reaction of anomeric OR groups is their displacement by X groups to form glycosyl halides. The latter are used in the Koenigs–Knorr method of glycosidation, which involves the reaction of a glycosyl halide with an alcohol in the presence of a heavy-metal catalyst. For example, tetra-*O*-acetyl-α-D-glucopyranosyl bromide reacts with methanol in the presence of silver carbonate (which acts as an acid acceptor) to afford, in high yield, methyl tetra-*O*-acetyl-β-D-glucopyranoside. The reaction mechanism is presumed to involve an intermediate carbonium ion (see Scheme 13).

It should be noted that the direct replacement of a hemiacetal hydroxyl group by a halogen atom (without protection of the hydroxyl groups) is not practical from the synthetic point of view, because it is usually accompanied by extensive isomerization and aromatization (including the formation of furan derivatives).

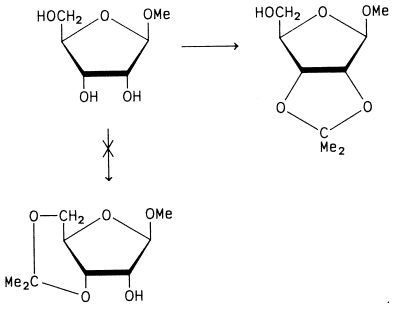


SCHEME 13 Mechanism of glycoside formation.

4. Nucleophilic Attacks by Hydroxyl Groups

There are two types of hydroxyl groups in the acyclic forms of monosaccharides: a primary hydroxyl group, which is, by definition, always terminal, and a number of secondary hydroxyl groups. In the cyclic forms of sugars, there exists a third type, namely, the glycosidic hydroxyl group attached to the anomeric carbon atom that forms part of the hemiacetal group of aldoses, or the hemiacetal group of ketoses. All three types of hydroxyl group are strong nucleophiles that can add to the carbonyl group of acid anhydrides, to afford esters, or replace the leaving groups of alkylating agents, to afford ethers. They can also induce substitution reactions at the anomeric center of the same or a different sugar, to afford anhydro derivatives or disaccharides, respectively. Finally, they can add in pairs to the carbonyl group of aldehydes and ketones, to gives cyclic acetals.

The hemiacetal hydroxyl groups of cyclic sugars are the most reactive of the three types of hydroxyl group. They are followed in nucleophilicity by the reactivity of the terminal primary hydroxyl groups. Because acyclic sugars do not possess hemiacetal hydroxyl groups, their primary hydroxyl groups are the most reactive hydroxyl groups in the molecule. The least reactive hydroxyl groups in both cyclic and acyclic sugars are the secondary hydroxyl groups. These may differ in reactivity, depending on whether they are axially or equatorially oriented (eliminations are favored by an antiperiplanar orientation) and according to whether the substituents are situated in a crowded environment. Thus, it is often possible to block a primary hydroxyl group and leave the secondary hydroxyl groups free by making use of the fact that, because secondary hydroxyl groups are in a more crowded environment, they might not react to any appreciable extent if mild reaction conditions and insufficient reagents are used. Stereo-chemical considerations also play important, and sometimes decisive, roles in determining the course of competing reactions involving more than one hydroxyl group. For example, when methyl β-D-ribofuranoside, which possesses one primary and two secondary hydroxyl groups, reacts with acetone, the 2,3-*O*-isopropylidene derivative is preferentially formed by attack of the two cis-oriented secondary hydroxyl groups, despite the fact that a primary hydroxyl group (which is trans-oriented) is available (see Scheme 14). This is because, had the latter group reacted, it would have produced a highly strained, six-membered 3,5-isopropylidene ring.



SCHEME 14 Formation of 2,3-*O*-ispropylidene rings.

The foregoing generalizations are useful when one is planning multistage syntheses of complex sugar molecules, especially when it is advantageous to block certain hydroxyl groups selectively and leave the other ones free for subsequent reaction. They are also useful in predicting which group will preferentially react when a limited amount of reagent is available.

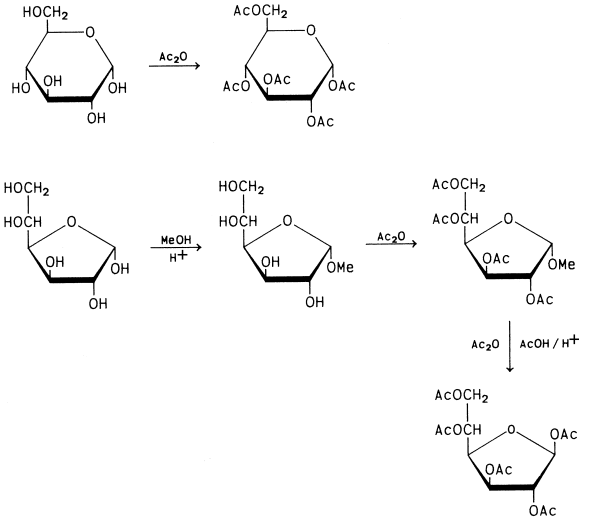
In the following section, four types of hydroxyl group derivatives will be discussed: esters, ethers, anhydro sugars and disaccharides, and cyclic acetals.

a. Formation of esters. Esters are generally used to block hydroxyl groups, that is, to deactivate their oxygen atoms and, by so doing, prevent them from attacking nucleophile acceptors. The esters most commonly used for this purpose are the acetates and benzoates. Occasionally, substituents are introduced in the para position of the phenyl ring of the latter esters in order to increase their crystallizing properties. The *O*-*p*-nitrobenzoyl and the *O*-*p*-toluoyl derivatives have been found to be useful in this respect.

Peracetylation (full acetylation) can be achieved at room temperature by treatment of the saccharide in pyridine with acetic anhydride or, at a higher temperature, by heating of the saccharide in a mixture of acetic acid and acetic anhydride. In both cases, the thermodynamically favored pyranose derivative is obtained. If the furanose derivative is desired, the methyl furanoside is acetylated, and the product is subjected to acetolysis (hydrolysis and acetylation), to replace the OMe group by OAc. The last reaction is conducted at low temperature with a mixture of acetic acid, acetic anhydride, and a few drops of sulfuric acid (see Scheme 15).

To prepare benzoates, *p*-substituted benzoates, and sulfonates, the necessary acid chloride is allowed to react in pyridine with the saccharides or saccharide derivatives.

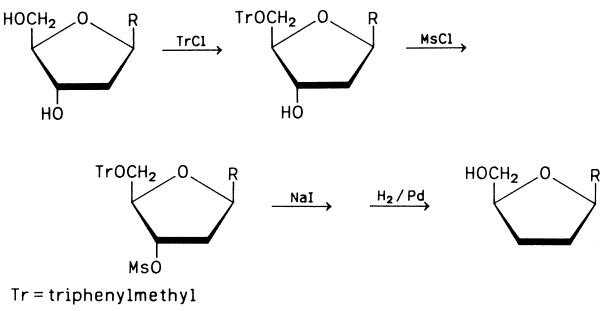
In general, ester groups are more stable in acidic than in basic media, and acetates are more readily hydrolyzed than benzoates. To carry out a deacetylation, a solution of sodium methoxide is added in catalytic amounts to



SCHEME 15 Formation of penta-*O*-acetyl-D-glucoyranose and furanose.

the sugar acetate at 0°C. Most esters are also saponifiable with NaOH in acetone at low temperature. Acetyl, benzoyl, and *p*-nitrobenzoyl groups attached to the oxygen atom of the anomeric centers of cyclic sugars can be displaced by nucleophiles, especially in the presence of Lewis acid catalysts. This reaction is useful in the preparation of glycosides and nucleosides. Ester groups attached to the other carbon atoms of a saccharide, that is, linked to nonanomeric carbon atoms, are considerably more inert toward nucleophiles and will not undergo substitution reactions under mild reaction conditions. To achieve a nucleophilic substitution of esters attached to nonanomeric centers, the latter esters must themselves be good leaving groups, for example, tosylates, mesylates, and triflates (see Scheme 16).

b. Formation of ethers. True saccharide ethers have the hydroxyl groups that are attached to the nonanomeric carbon atoms replaced by alkoxyl groups. These derivatives should be distinguished from the structurally similar



SCHEME 16 Replacement of an OH group with hydrogen.

alkyl glycosides, discussed earlier, which possess acetal-type OR groups. Methyl ethers have been used extensively in the structure elucidation of saccharides. They were first used by Haworth to determine the ring size of monosaccharides and the ring size and position of linkage of oligosaccharides. More recently, they have been used to volatilize monosaccharides before they are subjected to gas-chromotographic analysis. Because silyl derivatives are easier to prepare than methyl ethers, the former have now replaced methyl ethers in the gas-chromatographic analysis of monosaccharides. On the other hand, because methyl ethers are stable toward acid- and base-catalyzed hydrolysis, they continue to be used as a means of labeling free hydroxyl groups in saccharides, a procedure frequently used in the structure elucidation of oligosaccharides and polysaccharides. It is for this reason that the problem of permethylating saccharides (etherifying all of their hydroxyl groups) continues to attract the attention of carbohydrate chemists. The original methods of Purdie (MeI and AgOH) and of Haworth (Me_2SO_4 in alkali) have been much improved by the use of such aprotic solvents as $HCONMe_2$ (DMF) or Me_2SO (DMSO). Today, the most widely used methylating procedure is that of Hakomori, who used sodium hydride and Me_2SO to generate the base $MeSOCH_2^-$ needed for this type of methylation.

Benzyl ethers offer unique advantages in syntheses requiring the selective blocking and deblocking of hydroxyl groups. They can be introduced under mild conditions by the action of benzyl chloride in pyridine and can be removed in neutral media by catalytic hydrogenolysis, which does not affect esters or cyclic acetals. The former class of compounds is base labile, and the latter, acid labile.

Other synthetically useful ethers are the triphenylmethyl ethers, or as they are often called, trityl derivatives (Ph_3C—). Their bulky phenyl groups render their formation from secondary and tertiary hydroxyl groups so difficult that they are normally obtained from primary hydroxyl groups only. These ethers are therefore used to block the primary hydroxyl groups selectively, so as to leave the secondary hydroxyl groups free for subsequent reactions. Removal of the trityl ethers is very facile and can be achieved by mild acid hydrolysis. Such organic acids as acetic acid, which do not affect esters or cyclic acetals, selectively deblock the oxygen atom bearing the trityl group and regenerate the primary hydroxyl group, with liberation of triphenylmethanol. Alternatively, the trityl ether group can be removed by catalytic hydrogenolysis, which then affords triphenylmethane and the free primary hydroxyl group. The ease of hydrolysis of trityl ethers may sometimes be responsible for the occurrence of undesirable hydrolysis during the course of a subsequent reaction, particularly when vigorous conditions are used.

c. Anhydrides and disaccharides. An intra or intermolecular nucleophilic attack initiated by the oxygen atom of a suitably placed hydroxyl group on the anomeric carbon atom of the same or a different sugar will afford an anhydride or a disaccharide, respectively. The reaction is best performed by introducing a good leaving group, usually a halogen atom, on the anomeric carbon atom and blocking all of the hydroxyl groups except the one to be involved in the subsequent reaction (see Scheme 17). In this way, the formation of undesired products can be avoided.

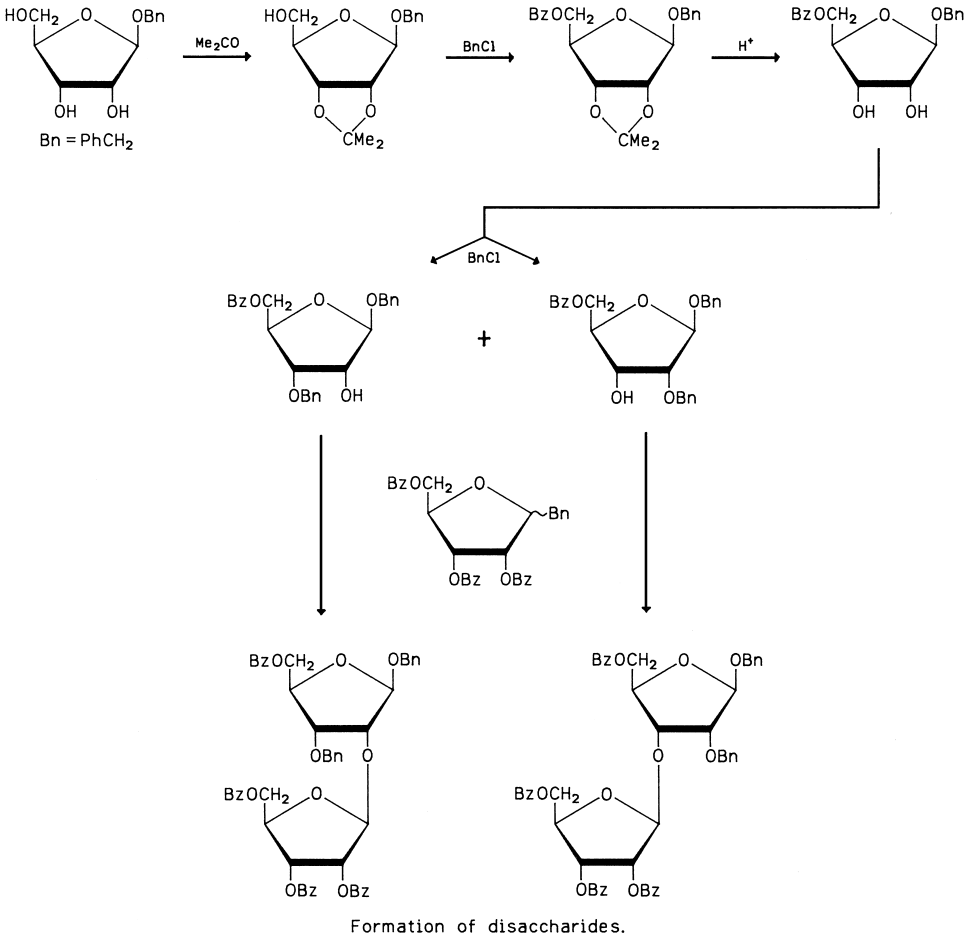
Disaccharides are often obtained enzymatically, by passing the substrate or substrates (the saccharides involved in the dimerization) through a column filled with a polymer that had been prebound chemically to the specific enzyme needed to perform the transformation. This method has, of late, acquired attention in industrial operations, because it allows the enzymes to remain active indefinitely on the column and prevents their being washed away during the work-up.

d. Cyclic acetals. If oriented properly, any two adjacent (but not necessarily contiguous) hydroxyl groups will react with an appropriate aldehyde or ketone to yield an unstrained five- or six-membered cyclic acetal, respectively. The most commonly used carbonyl compounds are benzaldehyde, which affords mostly six-membered benzylidene acetals, and acetone, which yields five-membered isopropylidene acetals. The latter derivatives have the advantage of existing in one isomeric form, unlike benzylidene derivatives, which exist in two isomeric forms. This is because, on reacting with chiral glycols, all aldehydes, except formaldehyde, and all mixed ketones yield chiral acetals.

The formation of cyclic acetals is catalyzed by acids and proceeds by two successive nucleophilic attacks; at first one hydroxyl group attacks the protonated carbonyl derivative to form the hemiacetal. The latter, in turn, becomes protonated and is attacked by the second hydroxyl group.

The formation of isopropylidene acetals can be achieved by treatment of a saccharide derivative with acetone or its dimethyl acetal (2,2-dimethoxypropane) in the presence of an acid catalyst such as HCl or H_2SO_4 and a dehydrating agent. The reaction is often conducted at room temperature and can be monitored by ^{1}H NMR spectroscopy. The signals of two methyl groups are usually well resolved because of their chiral environment.

A useful starting material in many syntheses is 1,2 : 5,6-di-*O*-isopropylidene-α-D-glucofuranose, obtained by treating D-glucose with acetone. This derivative affords an easy means of access to furanoses having a free hydroxyl group on C-3.



Formation of disaccharides.

SCHEME 17 Formation of disaccharides.

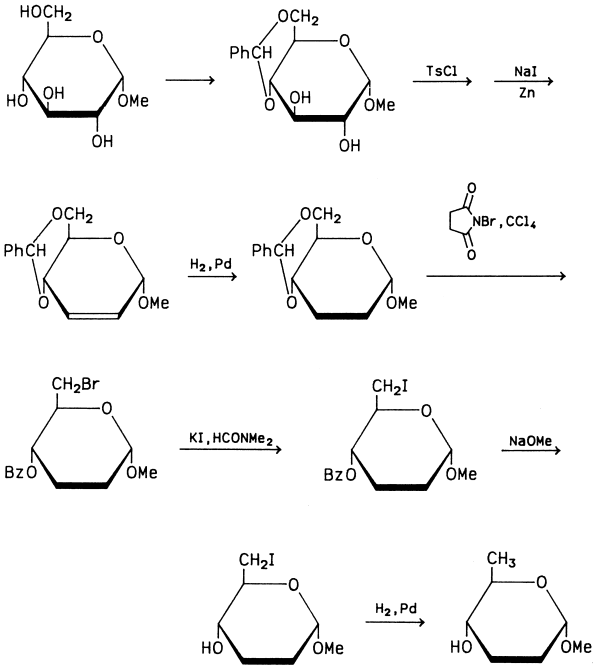
Benzylidene groups are also useful in the selective blocking of saccharide derivatives. They usually involve the primary hydroxyl group of a pyranoside and the 4-hydroxyl group if it is suitably positioned. Benzylidene groups are also useful for introducing a halogen atom instead of the primary hydroxyl group by the Hanessian method using *N*-bromosuccinimide as shown in Scheme 18.

5. Oxidation

The oxidation of a sugar involves the breaking of C—H, C—C, O—H bonds and the transfer of electrons to the oxidant. Two types of oxidation are possible. The first involves the transfer of two electrons from one atom to the oxidant and is referred to as heterotytic oxidation. The other, known as homolytic oxidation, occurs in two steps, each involving the transfer of one electron.

a. Heterolytic oxidations. Heterolytic oxidations, that is, those involving two-electron transfers, are exemplified by the oxidation of a secondary hydroxyl group to a keto group. This involves the elimination of two hydrogen atoms and the formation of a double bond between the carbon atom and the oxygen atom of the alcohol function. The first step takes place by breakage of a C—H bond and elimination of a hydride ion. Because the hydride anion is a very poor leaving group, this bond breakage is generally the slowest step of the oxidation (the rate-determining step). What makes the reaction possible, despite this, is the fact that the elimination is aided by the oxidant, which captures the pair of electrons and converts the hydride ion to a proton. Another proton is concurrently liberated by dissociation of the O—H bond of the alcohol, and both protons react with a base (usually a hydroxyl group to form water). Such an oxidation usually occurs by a concerted or E2 type of mechanism, in which the oxidant is not bonded to the sugar, except possibly in the transition state.

Oxidants, for example, chromic acid, have been shown to form intermediate esters, which subsequently decompose by bimolecular eliminations. The difference between this and the previous oxidation is that the leaving group is



SCHEME 18 Replacement of three OH groups with hydrogen.

the reduced form of the oxidant, and the capture of electrons by the oxidant is the driving force of the reaction. Furthermore, the breaking of the C—H bond, which occurs simultaneously, is the rate-determining step.

Of particular interest is the oxidation of primary hydroxyl groups with pyridinium chlorochromate (PCC) to afford aldehydes. The oxidant may be generated *in situ* by dissolving chromium trioxide in HCl to form chlorochromic acid and adding pyridine to form the desired salt. The oxidation is usually carried out in dichloromethane.

Glycols are more acidic than monohydric alcohols, and the C-1 group on an aldopyranose is even more acidic than either, because of the inductive effect of the ring oxygen atom. It is for this reason that sugars and glycosides are more readily oxidized than ordinary alcohols. Oxidation of free sugars at higher pH is often accompanied by competing processes of epimerization and degradation. In general, the β anomers of D sugars and their glycosides are more rapidly oxidized than the α anomers. Similarly, the 2-hydroxyl group of methyl β-D-glucopyranoside is more acidic than the corresponding group in the α-D anomer.

Halogens (usually bromine) and hypohalites (particularly sodium hypoiodite) have been used to oxidize aldoses to aldonic acids and their lactones. The aldehydic group of acyclic sugars is converted to a carboxyl group, and the anomeric hydroxyl group is converted to lactones. Although the terminal (primary) hydroxyl groups of sugars can also undergo oxidation by these reagents, the reaction occurs at a somewhat lower rate, so that the oxidation can be stopped at the monocarboxylic acid (or lactone) stage. By the use of moderately vigorous conditions, oxidations of the primary hydroxyl groups have been used to convert glycosides to glycosiduronic acids. The oxidation of secondary hydroxyl groups to keto groups requires drastic conditions, and these may be accompanied by cleavage of C—C bonds. A study of the above-mentioned oxidations has revealed that changes in the reaction conditions (temperature, acidity, and concentration) are often accompanied by changes in the nature of the oxidizing species. For example, the concentration of hypohalous acid is very low, but when sufficient alkali is added to the system, the concentration of hypohalite ion increases dramatically (at pH 1, 82% of the total chlorine exists as free chlorine and 18% as hypochlorous acid; at pH 4, only 0.4% of the chlorine is free and 99.6% exists as hypochlorous acid; and at pH 8, 21% exists as hypochlorous acid and 79% as hypochlorite). The order of effectiveness of halogens as oxidants is $Br_2 > Cl_2 > I_2$, the last being quite ineffective. This order corresponds to the rate of hydrolysis of the free halogen by water and to the solubility in water, which explains the ineffectiveness of free iodine as an oxidant. The mechanism of the oxidation of aldoses with chlorine and bromine in acid media was studied by Isbell, who found that the active oxidants are the halogens and not the hypohalous acids.

The accumulation of hydrogen bromide during oxidations by bromine profoundly lowers the rate of further oxidation. This effect is not due only to the simple increase in acidity, because, although other strong acids also inhibit the rate, the effect is largest for hydrogen bromide and chloride.

To minimize this inhibiting influence, the reaction can be conducted in the presence of a solid buffer, such as barium carbonate. In general, the presence of a buffer increases the yield of aldonic acid and precludes the hydrolysis of disaccharides. Yields of 96% of D-gluconic acid have been reported. When the oxidation period is extended for unbuffered solutions, keto acids may be formed in small yields. Thus, hexoses afford hex-5-ulosonic acids. Under more drastic conditions, C—C bonds are cleaved, yielding chain-shortened acids. The commercial production of aldonic acids has been achieved by electroysis of dilute solutions of sugars in the presence of a bromide and a solid buffer, such as calcium carbonate. Presumably, there is formed, at the anode, free bromine, which then oxidizes the aldose to the aldonic acid and is itself reduced to bromide. If the electrolytic method is not well controlled, aldaric acids and glyc-2 and -5-ulosonic acids may be produced.

Ketoses are generally resistant to bromine oxidation, and the latter may be useful in the removal of aldose contaminants from ketoses. However, by extending the period of oxidation, D-fructose affords D-*lyxo*-5-hexulosonic acid.

b. Homolytic oxidations. Most homolytic oxidations occur in two successive steps, each of which involves a one-electron transfer. The most difficult (and slowest) step in the homolytic oxidation of a sugar is the first, which involves the abstraction of a hydrogen atom from a C—H group, followed by the transfer of an electron from this hydrogen atom to the oxidant, to afford a proton. In this process, the sugar that lost a hydrogen atom is converted to a radical, which is stabilized by resonance (the unshared pair of electrons on the adjacent oxygen atom can migrate to the carbon atom from which hydrogen had been abstracted). This explains why the initial abstraction of hydrogen usually takes place by the removal of a hydrogen atom attached to carbon (H—|—C—O—H), rather than from the one attached to oxygen (H—C—O—|—H). It is also in agreement with the fact that compounds having equatorial hydrogen atoms linked to carbon atoms (i.e., carbon-linked hydrogen atoms, which are more accessible to the oxidant than the axially oriented hydroxyl groups attached to the same carbon atom) undergo more facile catalytic oxidation than axially oriented C—H groups (which are linked to equatorial hydroxyl groups). The final step of the oxidation is a fast reaction involving the homolysis of the O—H bond of the sugar radical, to afford a carbonyl group plus a second hydrogen radical which, like the first, is converted by the oxidant to a proton.

Degradation of a carbohydrate having a carbonyl group or potential carbonyl group begins with nucleophilic addition of hydroperoxide anion to the group, forming a hydroperoxide hydrin called the hydroperoxide adduct.

The adducts of substances having a hydroxyl group attached to the carbon atom adjacent to the carbonyl group usually decompose by a process that is called the β-hydroxy hydroperoxide cleavage reaction. With aldoses, this affords formic acid and the next lower aldose.

α,β-Diketones react rapidly with alkaline hydrogen peroxide with rupture of the —C—C bond and production of the two carboxyl groups. Degradation of an α,β-unsaturated ketone takes place by addition of hydroperoxide anion to the double bond. The adduct decomposes, producing an epoxide which may react further, as in the oxidation of the enolic form of diketomyoinositol.

The degradation of a reducing disaccharide proceeds relatively rapidly by the β-hydroxy hydroperoxide cleavage reaction until the process is interrupted by the glycosidic linkage. At this point, the hydroperoxide adduct, lacking an α-hydroxyl group, decomposes largely by elimination of the C-1 hydrogen atom, cleavage of the O—H bond, and formation of the corresponding 2-*O*-glycosylaldonic acid or by a minor reaction involving cleavage of the C-1—C-2 bond with formation of the *O*-formyl acetal of the next lower sugar.

6. Reduction of Carbohydrates

Numerous reagents have been employed to reduce carbohydrates and their derivatives. These include such inorganic hydrides as those of aluminum and boron; reactive metals, for example, sodium in the presence of proton donors; and metal catalysts, such as palladium, platinum, or nickel in the presence of molecular hydrogen.

a. Lithium aluminum hydride. Reductions of carbohydrate derivatives with lithium aluminum hydride are usually performed in nonaqueous solutions (such as benzene, ether, or tetrahydrofuran), because the reagent reacts violently with water. The boric acid formed during the borohydride reduction is usually removed as volatile methyl borate. Lithium aluminum hydride converts carbonyl groups to hydroxyl groups and ether lactones to alditols. For example, 2,3,5,6-tetra-*O*-acetyl-D-glucono-1,4-lactone yields D-glucitol. (The acetyl groups are reductively cleaved by the reagent.)

If less reactive hydrides are desired, to avoid reducing the aldehydes first formed to hydroxyls, then lithium tri-*ter*-butoxyaluminum hydride or diisobutylaluminum hydride (DIBAL-H) may be used at low temperature. The first will reduce an aldonic acid chloride into an aldose and the second an ester or a nitrile to an aldehyde (Scheme 19).

b. Borohydrides and diborane. In contrast to lithium aluminum hydride, sodium borohydride and lithium borohydride are relatively stable in water and can be used to reduce lactones to sugars in acid media or to the alditol in basic media. Thus, at pH 5, D-glucono-1,5-lactone affords D-glucose and, at pH 9, D-glucitol.

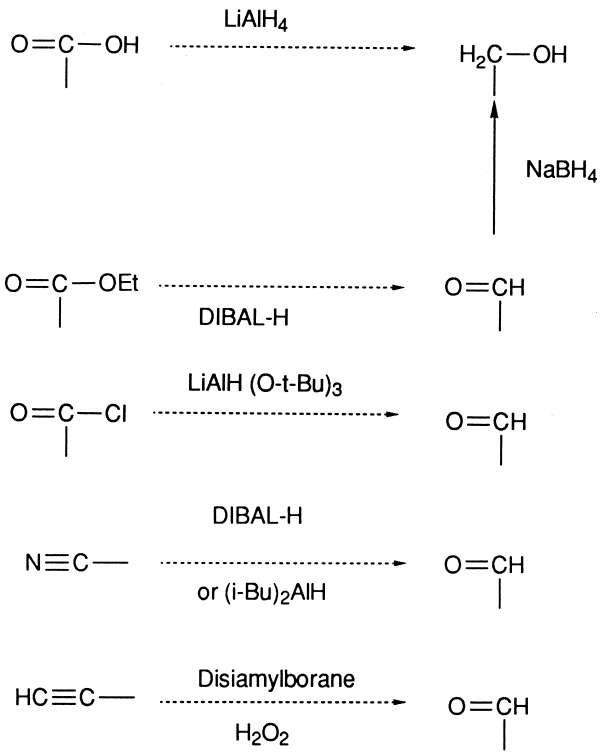
Sodium borohydride (or sodium borodeuteride) is also used to convert aldoses or oligosacchrides to alditol derivatives for examination by gas–liquid chromatography–mass spectrometry. Unlike hydride ions, which are electron-rich nucleophiles, diborane is the dimer of an electron-deficient electrophilic species (BH_3). It is formed by reaction of a borohydride ion with strong acids and is used in 2.2′-dimethoxydiethyl ether (diglyme) as the solvent. Diborane reacts faster with carboxylic acids, reducing them to primary alcohols, than with esters, because the carbonyl character of the adduct of the first is stabilized by resonance, and that of the second is destabilized by resonance. Diisoamylborane, which does not cleave ester groups (as does diborane), reduces acetylated

aldono-1,4-lactones to the corresponding acetylated aldofuranoses.

c. Sodium amalgam. Sodium amalgam in water and sodium in liquid ammonia are commonly used as reducing agents. The metal alloy has an electron-rich surface that reacts with protons of a proton-donating solvent to form hydrogen atoms, which may combine to form molecular hydrogen or react with the substrate to give a radical ion. The same radical-ion can be directly obtained from the metal, which, in both cases, is oxidized to a cation, giving up electrons to the substrate. The hydrogen radicals and the substrate radical-ions, which are bound by adsorption to the electron-rich surface of the metal, combine to afford the alkoxide ion, which then accepts a proton from the solution. In the early literature, the reduction of lactones to aldoses (a key step in ascending the series) had been carried out with sodium amalgam in acid solution (pH 3), using such buffers as sodium hydrogenoxalate, sulfuric acid, and, later, ion-exchange resins. These buffers are needed because the reduction does not proceed in alkaline solution (formation of the sodium salt of the aldonic acid impedes reaction with the nucleophilic reductant). The free acids are not reduced, but esters can be. Thus, methyl D-arabinonate is converted to D-arabinose by sodium amalgam. The susceptibility of lactones to reduction and the inertness of the free acids have been exploited in the preparation of glycuronic acids by reduction of the monolactones of aldaric acids.

d. Catalytic hydrogenation. In the course of hydrogenation with palladium or platinum catalysts, surface reactions occur between such electron-deficient species as the double bonds of carbonyl groups (which are attracted to the surface of the catalyst) and the hydrogen atoms generated by splitting hydrogen molecules. Catalytic hydrogenation is used to convert keto groups to secondary alcohol groups. Thus, Raney nickel reacts with D-*xylo*-5-hexulosonic acid to give both D-gluconic and L-idonic acid. Usually, the axial alcohol is the preponderant product, because the reductant approaches the substrate from the less hindered side of the carbonyl group, to form an equatorial C—H bond. Reduction of carbonyl groups can be carried beyond the alcohol stage, to the deoxy stage. For example, methyl β-D-*ribo*-hexopyranosid-3-ulose is converted to the 3-deoxy derivative by hydrogen in the presence of platinum, and 1-deoxyalditols can be obtained from aldoses by the reduction of their dialkyl dithioacetals with Raney nickel.

The stereo specific *syn* hydrogenation of unsaturated sugars (glycals and acetylenes) may be carried out heterogeneously, using Lindlar catalyst ($Pd/CaCO_3$) conditioned in quinoline, or homogeneously using Wilkinson's catalyst. $Rh[(Ph)_3P]_3Cl$. Alternatively, *trans* addition is achieved by reduction with Li in $EtNH_2S$ at low temperature.



SCHEME 19 Action of reducing agents.

III. OLIGOSACCHARIDES

Oligosaccharides are polymeric saccharides that have, as their name denotes, a low degree of polymerization, DP (*oligo* means "few" in Greek). They are composed of 2 to 10 glycosidically linked monosaccharides, which can be liberated by depolymerization (e.g., by acid hydrolysis). Oligosaccharides having DPs of 2 to 3 are sweet tasting and are included among sugars, whereas higher members are devoid of taste and are not referred to as such.

Sucrose is the world's most widely used sweetener and the one produced in largest quantities. Its industrial preparation from sugar cane and from beets is now well established. The process starts with pressing and concentration of the juice. This is followed by precipitation of sucrose as a calcium complex with lime and its regeneration with SO_2. Refining is achieved by charcoaling and recrystallization. Recent innovations in the process includes shifting to continuous flow operations, which have considerably increased the efficiency of production. Up until recently the only competitor of sucrose was glucose syrup produced by acid hydrolysis, or enzymatic hydrolysis of

corn starch. Now three additional products produced from starch have entered the market and lowered the consumption of sucrose. They are fructose syrup, very high fructose syrup, and crystalline fructose. As a result, sucrose can account for only half the sweeteners consumed in the United States. Oligosaccharides are grouped into simple (or true) oligosaccharides, which yield on deploymerization monosaccharides *only*; and conjugate oligosaccharides, which are linked to nonsaccharides such as lipids and afford on deploymerization monosaccharides and aglycons. The simple oligosaccharides are further classified (1) according to DP, into disaccharides, trisaccharides, tetrasaccharides, and so on; (2) according to whether they are composed of one or more types of monosaccharides, into homo- and heterooligosaccharides; and (3) according to whether they do or do not possess a hemiacetal function at one terminus of the molecule, into reducing and nonreducing oligosaccharides.

Related homooligosaccharides can form homologous series; a *homologous series* of oligosaccharides is a group of similarly linked oligosaccharides that are composed of the same monomer and whose DP increases in the series one unit at a time. When homopolysaccharides are partially hydrolyzed, they often afford homologous series of oligosaccharides. For example, the maltooligosaccharides obtained by partial hydrolysis of starch comprise dimers, trimers, tetramers, and so on composed of α-D-glucopyranose units linked by $1 \rightarrow 4$ acetal bonds.

A. Structure

The complete structure of oligosaccharides is established when the following points are determined:

1. The DP, that is, the number of monosaccharide units present in the oligomer molecule
2. The nature of the monosaccharide monomer(s)
3. In the case of heterooligosaccharides, the monosaccharide sequence
4. The ring size (pyranose or furanose) and the position of linkage of the different monosaccharides ($1 \rightarrow$?)
5. The anomeric configuration (α or β)
6. The conformation of the monosaccharide rings

1. Determination of the Degree of Polymerization

The following procedures are recommended for determining the DP of oligosaccharides:

a. Mass spectrometry. If the molecular weight of the oligosaccharide is less than 1000, mass spectrometry can be used to determine its molecular weight and hence its DP. It must be remembered, however, that mass spectra produced by electron impact seldom show the molecular ions of saccharides and that other ionization methods (e.g., chemical ionization) must be used to reveal these ions. The silyl ethers of oligosaccharides or of their alditols (obtained by reduction and silylation) are often used for mass spectrometric measurements, since they are more volatile than free oligosaccharides. Alternatively, peracetylated disaccharides or esters of disaccharides or of their glycosides can be used in the chemical ionization mode. Figure 12 shows three mass spectra of a disaccharide glycoside obtained by negative chemical ionization (a), by electron impact (b), and by positive chemical ionization (c). Only the chemical ionization mass spectra (negative and positive) revealed molecular ions.

b. Chromatography. It is often possible to determine the DP of members of a homologous series of polymeric saccharides from their position on chromatograms relative to a known member of the series. For example, partial hydrolysis of starch affords D-glucose and a homologous series of maltooligosaccharides, consisting of di-, tri-, tetra-, pentasaccharides, and so on. This mixture separates on chromatographs according to DP. On paper chromatographs and on high-performance liquid chromatographs, the mobility of the saccharides is inversely proportional to their DP (i.e., the largest oligomers move slowest), and on gel filtration chromatography, it is directly proportional to DP (i.e., the largest oligomers move fastest). The fact that the oligomers are eluted in the order of increasing or decreasing DP, makes it possible to determine their DP by determining their order of elution relative to a known member. Thus, by recognizing maltose (and glucose) in chromatograms of partially hydrolyzed starch (Fig. 13), it is possible to identify the subsequent bands as maltotriose, maltotetrose, maltopentose, and so on.

c. Reducing power. The DP of reducing oligosaccharides, particularly those with low molecular weight, can be determined from their reducing power. For example, the reducing power of a maltooligosaccharide relative to glucose (taken as 100%) is matched with the calculated values for oligomers having different DPs. Thus, a disaccharide would be expected to have about half the reducing power of glucose (actually 53%), a trisaccharide, a third (actually 35%), and a tetrasaccharide, a fourth (26%) that of glucose, and so on. Accordingly, if a maltooligosaccharide exhibits a reducing power equal to 35.8% that of D-glucose, it can be safely assumed that it is a trisaccharide. The differences in the reducing powers of successive members of a homologous series of oligosaccharides decrease with increasing DP, so that beyond a DP of 5, the differences become too small for reliable DP measurements.

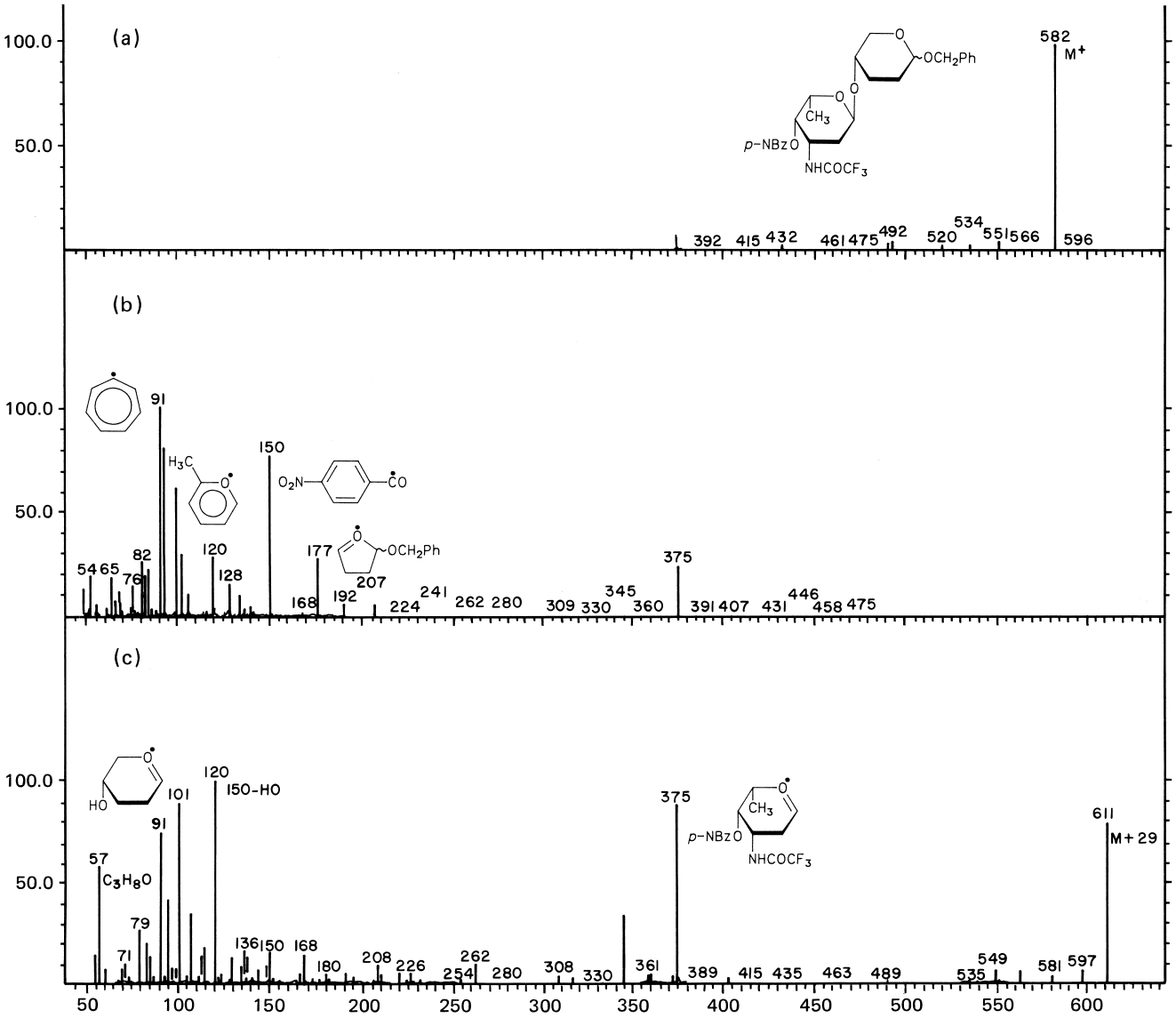


FIGURE 12 Mass spectra of a disaccharide. (a) Negative-ion chemical ionization in methane, (b) electron impact ionization, and (c) positive-ion chemical ionization in methane.

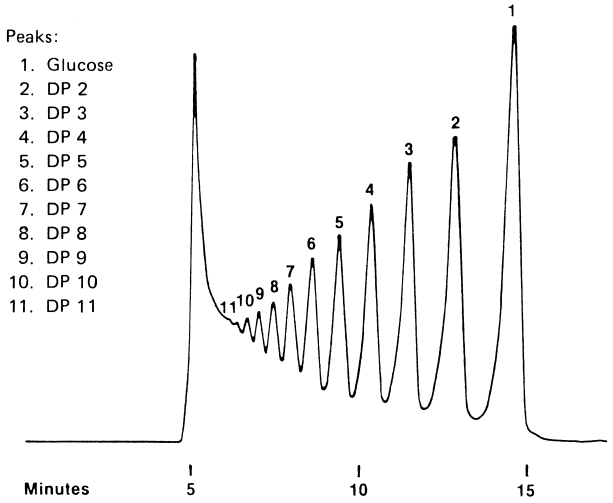


FIGURE 13 Liquid chromatogram of partially hydrolyzed starch showing glucose and the maltooligosaccharides.

2. The Monosaccharide Components

The nature of the monomers in an oligosaccharide is established by identifying the monosaccharides liberated by acid-catalyzed hydrolysis. Homooligosaccharides, which are composed of one type of monosaccharides, afford only one monosaccharide, which can be isolated from the hydrolysate by conversion to crystalline derivatives. On the other hand, heterooligosaccharides afford, on hydrolysis, a mixture of monosaccharides, which must be separated by chromatography. Paper, thin-layer, or high-performance liquid chromatography (HPLC) can be used without pretreating the monosaccharides, but a gas-chromatographic separation requires prior silylation of the saccharides, to volatilize them. It should be noted that silylated monosaccharides appear as double peaks (one peak for the α anomer and one for the β), which tends to crowd the chromatographs. To avoid this complication, the monosaccharides can be reduced before silylation (silylated alditols appear as single peaks). It should also be noted that all the above-mentioned chromatographic techniques cannot differentiate between D and L isomers, which mandates that polarimetric measurements (optical rotation, optical rotatory dispersion or circular dichroism) of the monosaccharide component(s) or of their derivatives be included in the structure elucidation of new oligosaccharides. In the case of heterooligosaccharides, this necessitates preparative chromatography of the hydrolysate before the optical measurements are taken.

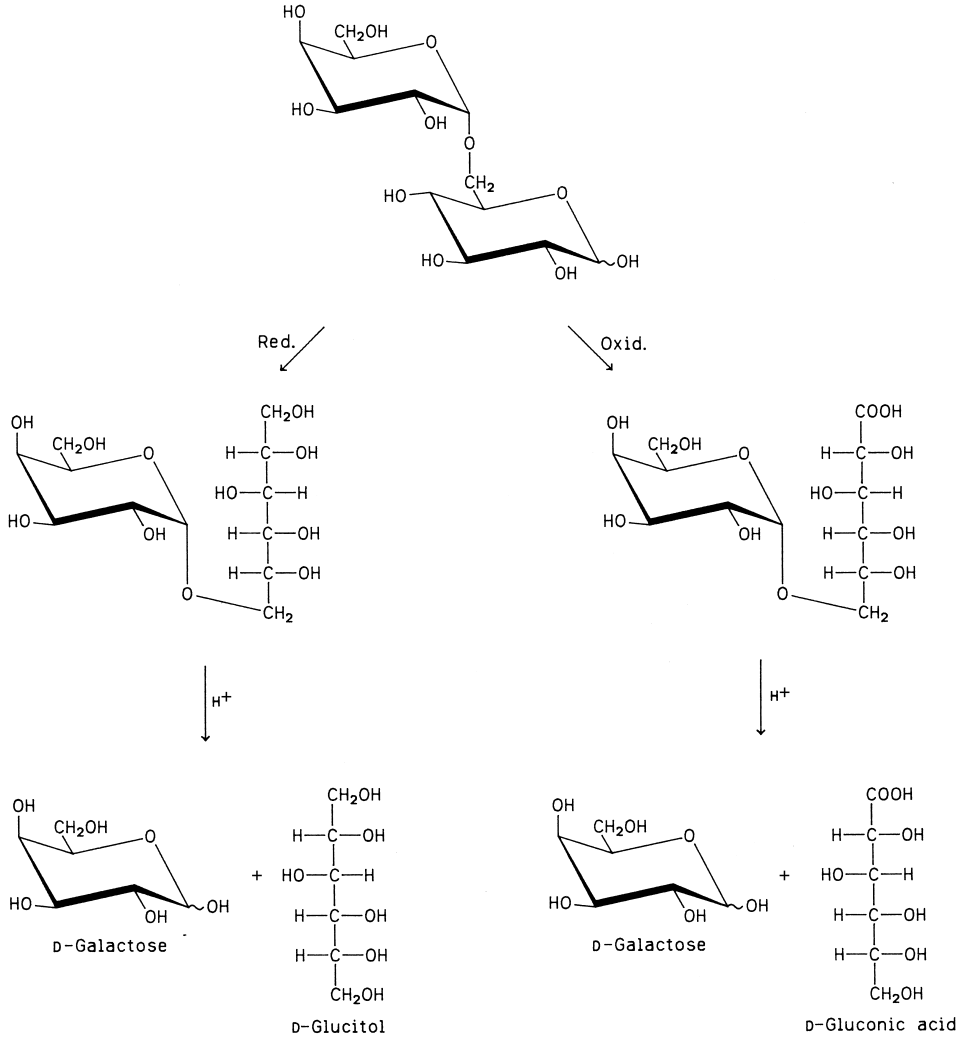
3. The Monosaccharide Sequence

Homooligosaccharides (whether reducing or nonreducing) do not require a monosaccharide sequence determination, because they possess identical monomers all through their chains. Heterooligosaccharides, except nonreducing disaccharides, must have the sequence of their monomers determined. In the case of reducing heterodisaccharides, it must be determined because the monomer at the nonreducing end of the molecule is different from the one at the reducing terminus, and the position of both monomers must be determined. It was stated above that nonreducing disaccharides do not require a monosaccharide sequence determination. This is because the two monomers are located at two similar (nonreducing) ends of a chain that has no beginning (no reducing terminus). On the other hand, nonreducing heterooligosaccharides with DP > 2 are made up of a monosaccharide linked glycosidically to the hemiacetal function of a reducing oligosaccharide, the sequence of which must be determined.

The methods available for determining the monomer sequence are presented here, in order of increasing DP of the oligosaccharides, starting with disaccharides.

a. Monosaccharide sequence in disaccharides. The sequence of monomers in disaccharides is determined when the monomer located at the reducing terminus of the molecule is identified. This automatically determines the position of the other monomer (it must be located at the nonreducing end). To identify the monosaccharide located at the reducing terminus, use is made of the fact that this saccharide moiety exists in equilibrium with an acyclic form and is therefore much more susceptible to oxidants and reducing agents than the other moiety. Thus, mild oxidation converts reducing disaccharides to aldobionic acids, which on hydrolysis afford aldonic acids (from the reducing termini) and reducing monosaccharides (from the nonreducing ends). Reduction affords aldobiitols, which on hydrolysis yield alditols, from the reducing moieties and monosaccharides from the nonreducing moieties. The use of this method of structure determination is exemplified by the oxidation and reduction of melibiose [6-(α-D-galctopyranosyl)-D-glucopyranose] to give, in the first case, melibionic acid [6-(α-D-galactopyranosyl)-D-gluconic acid] and, in the second, melibiitol [6-(α-D-galactopyranosyl)-D-glucitol] (Scheme 20). Hydrolysis of these affords D-galactose from the nonreducing end of the molecule and D-gluconic acid in the first case and D-glucitol in the second (both formed from the reducing half of the molecule). This experiment indicates that D-galactose is the monomer located at the nonreducing end of the dimer and that D-glucose is the one at the reducing terminus.

b. Monosaccharide sequence in trisaccharides. Because it is easier to determine the structure of disaccharides than it is to investigate the structure of trisaccharides,



SCHEME 20 Oxidation and reduction of melibiose.

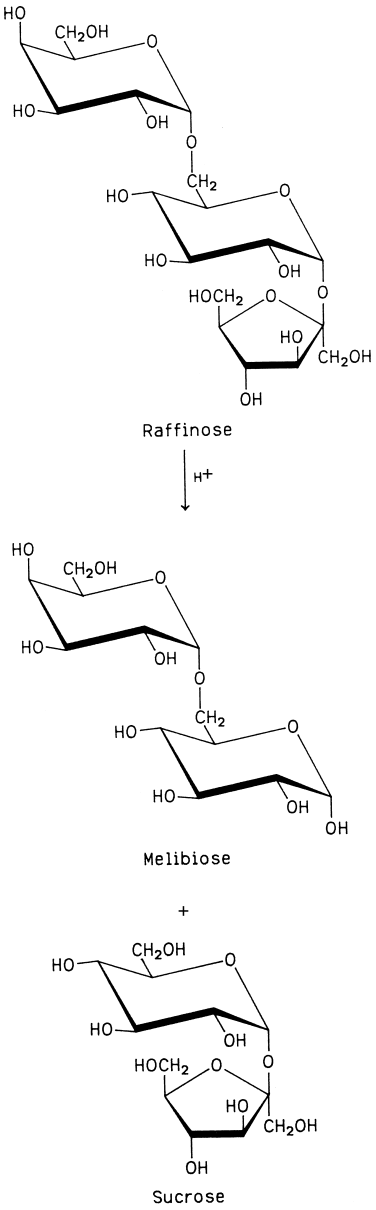
it is often advantageous to deduce the structure of the latter by identifying (or determining the structure of) the two disaccharides resulting from the partial hydrolysis (either chemical or enzymatic) of the trisaccharide under investigation. If known, the disaccharides would be compared with authentic samples, and if not, the new disaccharides would be investigated as shown above. In piecing together the structures of a trisaccharide from two disaccharides, the following rules are used:

1. If the trimer is composed of three (different) types of monomers (e.g., A–B–C), only one monomer, namely, B, will be found in both dimers (A–B and B–C). This monomer must be located in the center of the trimer, and the remaining monomers (A and C) are made to occupy the same terminal positions in the trimer as they do in the dimer. [The monomer present in the nonreducing end of a disaccharide (A) will be at the nonreducing end of the trisaccharide, and the monomer found at the reducing end of a disaccharide (C) will be put at the reducing end of the trisaccharide.]

2. If the trimer is composed of only two types of monomers (e.g., (A–A–B or A–B–B or A–B–A or B–A–B), it is possible to have one or two common monomers in the resulting dimers. If one monomer only is common to the two dimers, for example, A in A–A and A–B (obtained from A–A–B) or B in A–B and B–B (obtained from A–B–B), the same rules mentioned above are used. However, if both monomers (A and B) are common to the two dimers (e.g., A–B and B–A obtained from either A–B–A or B–A–B), an estimation (by gas chromatography or HPLC) of the two monomers in the trisaccharide

hydrolysate will reveal that one monomer (A in A–B–A and B in B–A–B) occurs in double the amount of the other; this monomer must be terminally located in the trisaccharide molecule. To illustrate this method, a nonreducing trisaccharide, raffinose, will be used as an example of a trisaccharide having three different monomers (A–B–C) (Scheme 21). On partial hydrolysis with acids, this trisaccharide affords a reducing disaccharide [6-(α-D-galactopyranosyl)-D-glucopyranose], and on enzymatic hydrolysis it yields a nonreducing disaccharide sucrose [α-D-glucopyranosyl-β-D-fructofuranose]. The common monomer in both disaccharides, D-glucopyranose, must therefore be located at the center of the trimer. The other two monomers (D-galactopyranose and D-fructofuranose) are then assigned positions at both ends of the nonreducing trimer. Oxidation (or reduction) of the first disaccharide, followed by hydrolysis, will reveal that D-glucose is at the reducing terminus and D-galactose is at the nonreducing ends of the dimer molecule and confirm the location of the latter at one of the nonreducing ends of the trimer molecule. Since the second disaccharide is nonreducing, the terminal D-fructofuranose must be attached through C-2 to C-1 of D-glucose. This establishes the structure of raffinose as *O*-α-D-galactopyranosyl-(1 → 6)-*O*-α-D-glucopyranosyl-(1 → 2) *O*-β-D-fructofuranoside.

c. Monosaccharide sequence in tetrasaccharides. The sequence in tetrasaccharides can be deduced by determining the sequence of monosaccharides in the oligosaccharides that result from their partial hydrolysis (two trisaccharides and three disaccharides). If the tetrasaccharide is composed of four different monomers (A–B–C–D), the monosaccharide sequence can be deduced by determining the monomer sequence in the disaccharides (A–B, B–C, and C–D) produced by partial hydrolysis. Use is made of the fact that the monomers common to two dimers (B, found in A–B and B–C; and C, in B–C and C–D) must be linked together and must be located in the center of the tetramer. (Thus, B must be attached to A and C, and C attached to B and D.) The same rule applies if only one monomer is repeated and the two analogous monomers are contiguous (as determined from the fact that *one* of the dimers is a homodisaccharide). Examples of such homodisaccharides are A–A obtained from A–A–B–C; B–B obtained from A–B–B–C; and C–C produced from A–B–C–C. If, on the other hand, the repeating monomers are not contiguous but are separated by one monomer (as in A–B–A–C and A–B–C–B) or by two monomers (as in A–B–C–A), a study of the dimers alone may not suffice to elucidate the structure of the tetramer, making it necessary to study also the structure of the two trisaccharides resulting from its partial hydrolysis.



SCHEME 21 Formation of melibiose and sucrose from raffinose.

In general, the structure elucidation of polymers becomes more difficult as the DP increases. However, because higher oligosaccharides are composed of repeating units seldom larger than tetrasaccharide fragments (and often composed of mono-, di- or trisaccharides), the number of possible oligosaccharides liberated on partial hydrolysis remains relatively small. Thus, whereas a tetrasaccharide (A–B–C–D) will afford on partial hydrolysis two trisaccharides (A–B–C and B–C–D) and three disaccharides (A–B, B–C, C–D), an octasaccharide (or for that matter a polysaccharide) composed of the same tetrasaccharide repeating unit will afford only two additional

trisaccharides (D–A–B and C–D–A) and one additional disaccharide (D–A).

4. Ring Size and Position of Linkage

The next stage in the structure elucidation of oligosaccharides involves the determination of the ring size and the position of linkage of the monosaccharide constituents. For nonreducing disaccharides, the position of linkage is by necessity between anomeric carbons; otherwise the disaccharide would be reducing. If the nonreducing disaccharide is composed of two aldoses, these two monomers must be linked by an acetal oxygen bridge joining C-1 of one aldoses to C-1 of the other, which is signified by (1 → 1); if the disaccharide is composed of two uloses (ketoses), the oxygen bridge must link C-2 of one ulose to C-2 of the other (2 → 2); finally, if the disaccharide is composed of one aldose and one ulose, the linkage is (1 → 2) (with the oxygen bridge linking C-1 of the aldose to C-2 of the ulose). Accordingly, when the monosaccharide components of a nonreducing disaccharide are identified, there is no uncertainty about the positions of linkage, but only about the size of the rings (and the anomeric configuration, which will be discussed later). Reducing disaccharides, on the other hand, must have their position of linkage and ring size determined. The position of linkage and/or ring size of the monosaccharide components of reducing and nonreducing oligosaccharides can be determined by labeling or by partial hydrolysis, as follows:

a. Labeling of free hydroxyl groups. The free hydroxyl groups in oligosaccharides are attached to carbon atoms that are not involved in ring formation or in glycosidic bonds. These carbon atoms can be recognized if they are marked with a suitable label, for example, by attaching to their oxygen atoms permanent blocking groups that will not be removed during the hydrolysis of the oligosaccharide. Methylation is often used, since many of the partially methylated monosaccharides that result from such hydrolysis have been characterized by gas chromatography. Methylation and hydrolysis of a reducing disaccharide afford two methylated monosaccharides. The first has two free hydroxyl groups (one at position 1 and one where the ring was attached), and the second possesses three positions free (one at position 1, one where the ring was attached, and one where the glycosidic bond was attached). The vacant positions in the latter might create ambiguity, since it is not known which of the unblocked positions were due to the ring and which to the glycosidic bond. To avoid this confusion, it is necessary to carry out another methylation on an acyclic disaccharide derivative, for example, the aldobionic acid obtained by oxidation of the disaccharide or the aldobitol obtained by its reduction. These compounds have one acyclic moiety (the aldonic acid and the alditol, respectively), so that after hydrolysis of the alkylated dimer, the unmethylated position in the acyclic moiety will be the position of the glycosidic linkage. If the disaccharide is nonreducing, the position of the glycosidic linkage is known (C-1 for an aldose and C-2 for an ulose), and only one methylation experiment is needed (the unlabeled position is where the ring was attached).

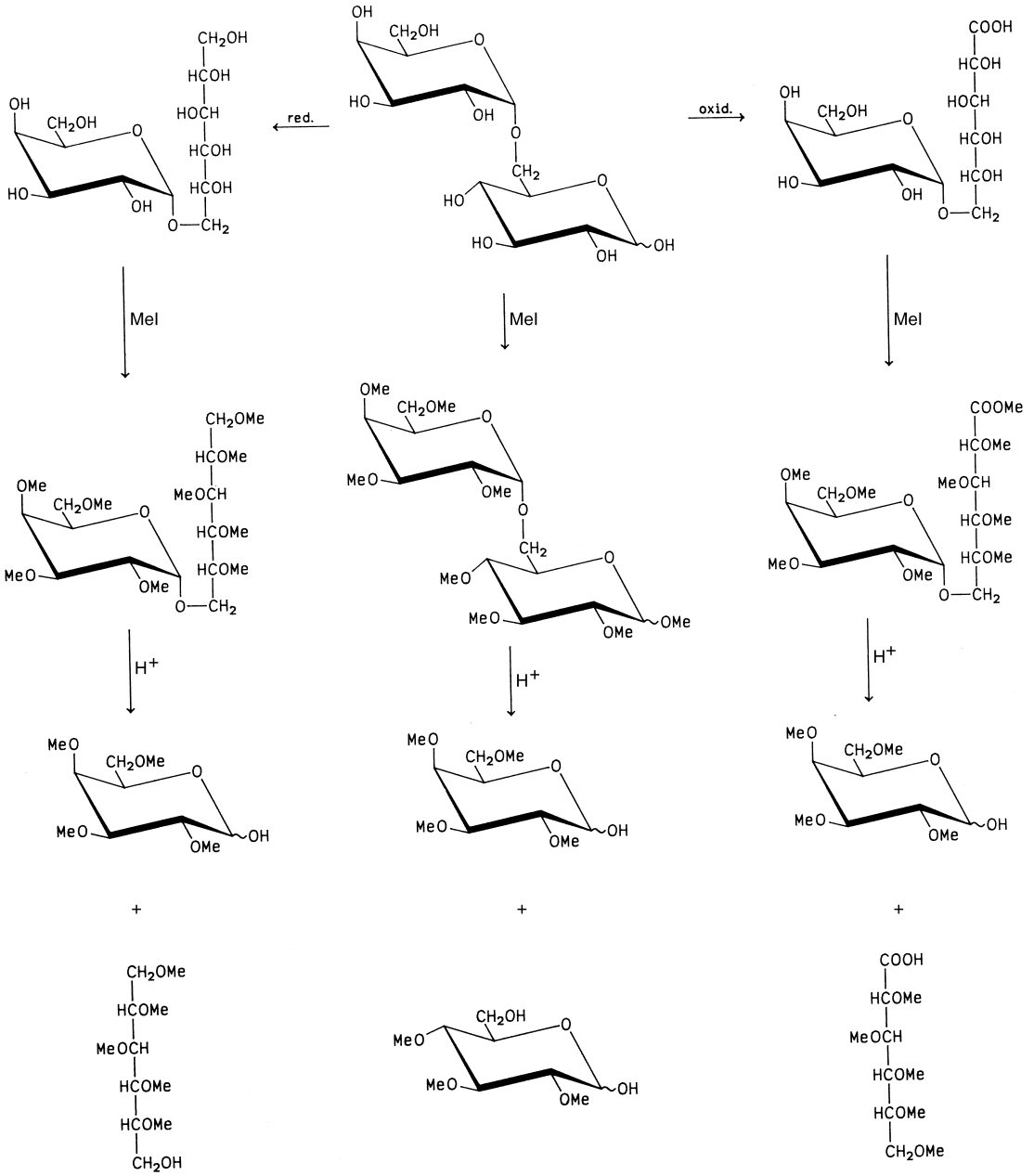
The use of labeling in structure elucidation is exemplified by the methylation and hydrolysis of melibiose to yield 2,3,4,6-tetra-*O*-methyl-D-galactose and 2,3,4-tri-*O*-methyl-D-glucose and by the oxidation of melibiitose to melibionic acid or its reduction to melibiitol, then methylation of these, followed by hydrolysis of the methylated aldobionic acid or aldobiitol to yield 2,3,4,6-tetra-*O*-methyl-D-galactose and 2,3,4-5-tetra-*O*-methyl-D-gluconic acid in the first case and 1,2,3,4,5-penta-*O*-methyl-D-glucitol in the second (see Scheme 22).

b. Partial hydrolysis. It is possible to use the structure of known disaccharides to determine the structure of higher oligosaccharides. Thus, the fact that the trisaccharide raffinose (discussed above) affords on partial hydrolysis melibiose and sucrose, whose structures are known to be 6-(α-D-galactopyranosyl)-D-glucopyranose and α-D-glucopyranosyl β-D-fructofuranoside, respectively, establishes that the trisaccharide molecule is composed of an α-D-galactopyranose ring attached through an α-glycosidic bond to position 6 of an α-D-glucopyranose ring, which in turn is attached glycosidically to the anomeric position of a β-D-fructofuranose ring. In other words, the trisaccharide must be 6-(α-D-galactopyranosyl)-α-D-glucopyranosyl β-D-fructofuranoside.

5. Anomeric Configuration

a. By partial hydrolysis. The hydrolysis of raffinose to melibiose and sucrose, discussed earlier, suggests that the linkage between the galactose and the glucose units in the trisaccharide is α-D, and the linkage between glucose and fructose is α-D for glucose and β-D for fructose.

b. By enzymatic hydrolysis. The anomeric configuration of the glycosidic bond of disaccharides can be determined by enzymatic hydrolysis. For example, emulsin, an enzyme obtained from bitter almonds, is known to hydrolyze β-D-glucosidic linkages and not α-D linkages. Accordingly, if a glucose-containing disaccharide is hydrolyzed by emulsin, it can be concluded that it possesses a β-D linkage. It is essential for structural work involving enzymes that the enzyme preparations be of the highest



SCHEME 22 Elucidation of the structure of melibiose.

purity in order to prevent misleading results caused by contaminants.

6. Conformation

a. By NMR spectroscopy. The most common method for determining both the anomeric configuration and the ring conformation of oligosaccharides is NMR spectroscopy. The coupling constant of the anomeric proton ($J_{1,2}$) is measured and used to determine the dihedral angle between H-1 and H-2 in the disaccharide. Since this angle depends not only on the anomeric configuration, but also on the ring conformation, both are determined concurrently. The sample is irradiated at the frequency of the H-1, so that the signal of H-2 is identified (since it is decoupled partly), and $J_{2,3}$ determined. The procedure is repeated by irradiating the sample at the frequency of the H-2 absorption (to identify the signal due to H-3 and measure $J_{3,4}$) and then at the frequency of H-3 to identify the signal due to H-4 and so on. When all the

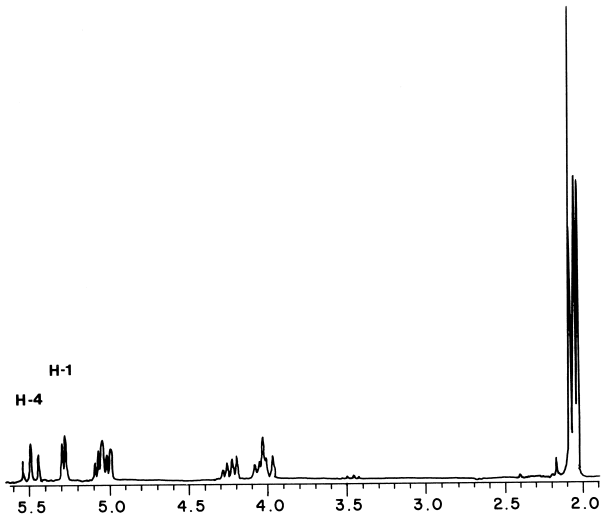


FIGURE 14 NMR spectrum of α,α-trehalose octaacetate.

coupling constants of one ring are measured, the process is repeated for the other ring. Finally, the Karplus equation is used to determine the dihedral angle between the different protons, which establishes the conformation of the ring and the anomeric configuration. To identify all the signals in an oligosaccharide spectrum, high-resolution NMR instruments are needed (preferably ones with two-dimensional mapping capabilities). In the absence of such equipment, it is still possible to determine the anomeric configuration and ring conformation by measuring the coupling constants of H-1 and H-4 (for pyranose rings). Figure 14 shows the NMR spectrum of octa-*O*-acetyl-α-D-glucopyranosyl-α-D-glucopyranoside (α,α-trehalose octaacetate), which clearly shows that this molecule possesses two identical α-D-glucopyranosyl rings in the $^1C_4$ conformation. This is apparent from the coupling of the anomeric proton and the fact that the two rings produce identical signals, as well as from the coupling of H-4 (split by the two trans-diaxial protons at H-3 and H-5).

b. By crystallography. Another way of determining the anomeric configuration and ring conformation is by crystallography (using either X-ray or neutron diffraction). Both techniques will afford the complete structure (in the solid state) of a crystalline oligosaccharide, including the orientation of the two rings vis-à-vis one another (angles ϕ and ψ). Figure 15 shows a diagram, deduced from X-ray defraction, of a dihydrate of α,α-trehalose. The anomeric configuration and the $^4C_1$ conformation of the two rings are clearly revealed. Note also the remarkable symmetry of the molecule, which agrees with the NMR data discussed earlier.

B. Oligosaccharide Synthesis

Three types of oligosaccharide synthesis will be discussed in this section; to choose between them one should consider the complexity of the oligosaccharide, the length of the proposed scheme, and the availability of its starting materials. The synthetic methods are

1. The standard chemical method, which usually involves a nucleophilic substitution of a leaving group, previously introduced on a mono- or oligosaccharide, by an appropriate saccharide adduct.
2. The same chemical reactions described in (1), but carried on a solid support. The operation resembles the automated synthesis of peptides and polynucleotides carried out fixed beds of polymer or resin. In the present case, the nonreducing end of the monosaccharide is linked to a resin, via a linker. The monosaccharide is then reacted, successively and in the desired order, with the different monosaccharide adducts. When the oligomer is formed, it is separated from the linker.
3. Biochemical methods which use one or more enzymes to carry out the oligosaccharide synthesis, either in solution or on a solid support.

1. Chemical Methods of Oligosaccharide Synthesis

The chemical synthesis of oligosaccharides usually involves reactions between saccharide derivatives that possess good leaving groups at the anomeric position, such as a halogen atom or an ester group, and an adduct, which may be a monosaccharide or an oligosaccharide. In the first

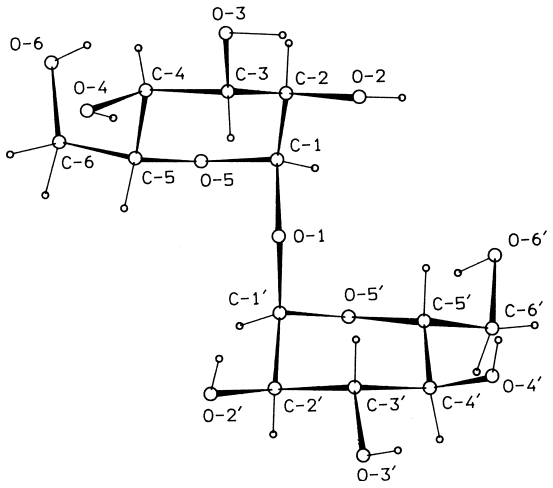


FIGURE 15 Conformation of α,α-trehalose dihydrate determined by X-ray crystallography.

case, the glycosyl halide is subjected to a Koenigs–Knorr type of reaction, with a halogen acceptor (e.g., $AgCO_3$) as catalyst, whereas when an ester is used, a Helfrich type of reaction is carried out with a Lewis acid catalyst, such as $SbCl_6$, BF_3, $SnCl_4$, TiF_4, etc. The glycosyl halides are best prepared from peracetylated or perbenzoylated mono- or disaccharides by treatment with dichloromethyl methylether (DCMME). Other good leaving groups are thioglycosides, which requires a promoter such as dimethyl(methylthio)sulfonium triflate (DMTST) and trichloroacetimidate which requires no activator and is therefore often used in oligosaccharide syntheses on fixed bed polymers, discussed below. The anomeric configuration of the formed glycosidic bond is controlled by the nature of the protecting group adjacent to the leaving group. Participating groups such the acetyl group lead to 1,2 trans glycosidic linkages and nonparticipating groups such as the benzyl group form 1,2-cis glycosidic linkages. In both cases the leaving group and the protecting group must be removed under mild conditions that do not affect the glycosidic linkages (bases in the first type and hydrogenation in the second).

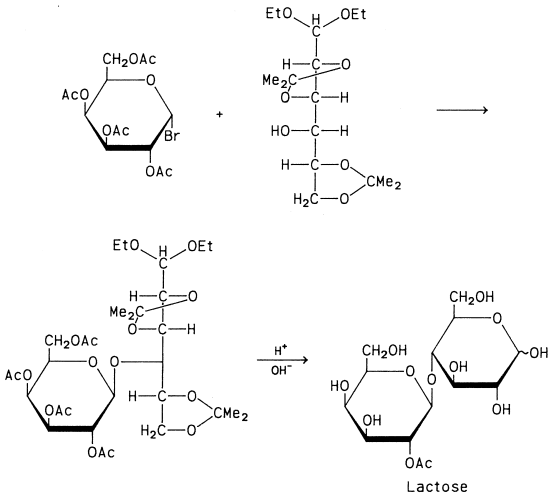
Whichever glycosidation method is used in oligosaccharide synthesis, two questions must be carefully addressed:

1. How can one ensure that only the oxygen atom in the desired position forms the glycosidic bond? This is achieved by ensuring either that the desired hydroxyl group is the most reactive hydroxyl group in the adduct molecule or, better, that it is the only one available for reaction. Because hemiacetal hydroxyl groups are the most reactive hydroxyl groups in cyclic saccharides, it is possible to prepare nonreducing disaccharides by reacting glycosyl halides with unprotected monosaccharides. On the other hand, to form reducing disaccharides, it is necessary to protect some or all of the nonreacting hydroxyl groups in the adduct. Thus, if the desired disaccharide is linked through a primary hydroxyl group [e.g., in the case of a (1→ 6)-linked disaccharide], it is necessary to block (by glycosidation) the anomeric position of the adduct (because the hemiacetal hydroxyl is more reactive than the primary hydroxyl group). Finally, if more than one primary hydroxyl group is present in the adduct (such as in the case of an ulose), or if one of the secondary hydroxyl groups is to form a glycosidic bond, it is advisable to block all but this hydroxyl group to ensure that only the desired glycosidic linkage is formed.

2. How can one ensure that the glycosidic linkage formed is of the desired anomeric configuration? The α and β configuration of a newly formed glycosidic bond is determined, to a large extent, by the blocking groups present in the sugar moiety undergoing nucleophilic attack. Thus, if a participating group (e.g. an acetyl or a benzoyl group) is attached to O-2 of the molecule undergoing nucleophilic attack, then a Koenigs–Knorr or Helfrich type of reaction will afford a glycoside having a trans-1.2 configuration. This isomer is favored because the intermediate cyclic carbonium ion is attacked from the side opposing the ring. This is why β isomers are obtained from D-glucopyranosyl halides (and D-galactopyranosyl halides) and α isomers are formed from D-mannopyranosyl halides.

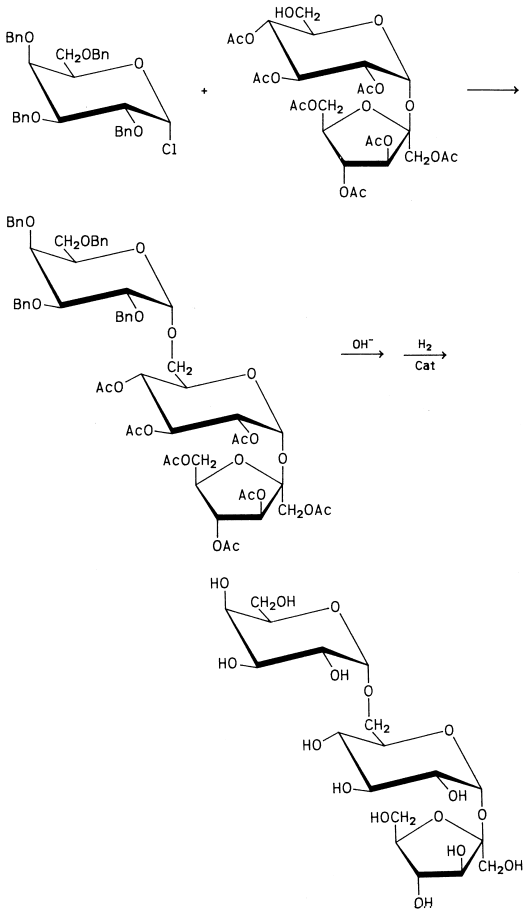
To obtain a cis-1,2 configuration (e.g., a glycoside having an α-D-glucopyranosyl, an α-D-galactopyranosyl, or a β-D-mannopyranosyl configuration), the OH-2 of the sugar moiety undergoing nucleophillic attack must be protected by a nonparticipating group, such as a benzyl group. In this case, a mixture of α and β isomers is obtained, which can be separated by chromatography. The composition of this mixture is influenced by anomeric effects, which favor α-D anomers, and by temperature and duration of reaction time, which favor the thermodynamically more stable product (i.e., the one with equatorial substitutents).

The following are examples of oligosaccharide syntheses that illustrate the ideas discussed above. Lactose [4-*O*-(β-D-galactopyranosyl)-D-glucose] possesses, as its name denotes, a β-D-glycosidic bond linking C-1 of galactose to position 4 of glucose. It could therefore be synthesized by reacting a galactopyranosyl halide having participating protecting groups on O-2 with a glucose derivative having all the hydroxyl groups blocked except for OH-4. Actually this synthesis was performed by reacting 2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl bromide with 2,3:5,6-di-*O*-isopropylidene-D-glucose diethylacetal under Koenigs–Knorr conditions, then hydrolyzing the acetal groups with acid and the ester groups with base (see Scheme 23).



SCHEME 23 Synthesis of lactose.

Consider now the synthesis of raffinose, which is *O*-α-D-galactopyranosyl-(1 $\rightarrow$ 6)-*O*-α-D-glucopyranosyl-(1$\rightarrow$2) β-D-fructofuranoside. It is evident that the galactopyranosyl halide needed for this synthesis must be protected by a nonparticipating group in order to afford the desired α-D-galactopyranosyl linkage (cis-1,2 configuration). Furthermore, since the adduct (sucrose) is a nonreducing disaccharide that possesses three primary hydroxyl groups (all of which are available for attack on the carbonium ion), it is necessary to block at least two of them to ensure that the desired primary hydroxyl group, namely, the one attached to C-6 of the glucopyranose moiety, is the one that reacts with the galactopyranosyl halide. Experimentally, raffinose was synthesized by reacting, under Koenigs–Knorr conditions, a benzyl-protected galactopyranosyl halide (tetra-*O*-benzyl-α-D-galactopyranosyl chloride) with a sucrose derivative (2,3,4,1′,3′,4′,6′-hepta-*O*-acetylsucrose) having ester groups replacing all hydroxyl groups except OH-6 of the glucopyranose moiety (see Scheme 24).



SCHEME 24 Synthesis of raffinose.

Although it is possible to achieve the synthesis of higher oligosaccharides by the addition of one monosaccharide at a time, it is advantageous to add the oligosaccharide components of a large oligomer in blocks of two or more monosaccharides (block synthesis). Thus, the trisaccharide and tetrasaccharide can be prepared by reacting a disaccharide glycosyl halide with a suitably protected mono- or disaccharide. It is evident that a second round of reactions would afford penta- and hexasaccharides.

Another interesting approach to the chemical and enzymatic synthesis of oligosaccharide is the use of a stationary polymer support to retain substrates (or enzymes) inside a reactor (usually a column), while successive reagents are passed through the polymer and are then washed out of the reactor. The advantage of this method is that it affords an efficient way of carrying out successive reactions without loss of the material attached to the column. For industrial enzymatic reactions involving costly enzymes, retaining these inside the reactor would seem quite attractive.

2. Synthesis of Oligosaccharides on Fixed Beds of Polymers

The importance of glycopeptides and glycolipids in medicine and the small amount of oligosaccharides released from natural sources necessitated the development of improved methods to prepare them including novel approaches in their synthesis. As in the case of automatic nucleotide- and peptide-synthesizers, a serious search has been made to develop suitable polymers for use as fixed beds and as linkers to attach saccharides and to facilitate the removal of the oligosaccharide formed at the end of the synthesis from the polymer support. The successive monomers are added one at a time to the linker until the desired oligomer is formed, then it is separated from the linker. Many oligosaccharide syntheses on polymer support use the commercially available polymer, polyethylene ω-monomethyl ether (MPEG) attached to the linker, α,α-dioxyxylyl diether (DOX). The product, (MPEG-DOX), needed for the oligosaccharide synthesis on solid support is prepared as follows:

$$\text{Me-O-(CH}_2\text{-CH}_2\text{)}_n\text{-CH}_2\text{-CH}_2\text{-OH} + \text{Cl-CH}_2\text{-(C}_6\text{H}_4\text{)-CH}_2\text{-CL} \rightarrow$$

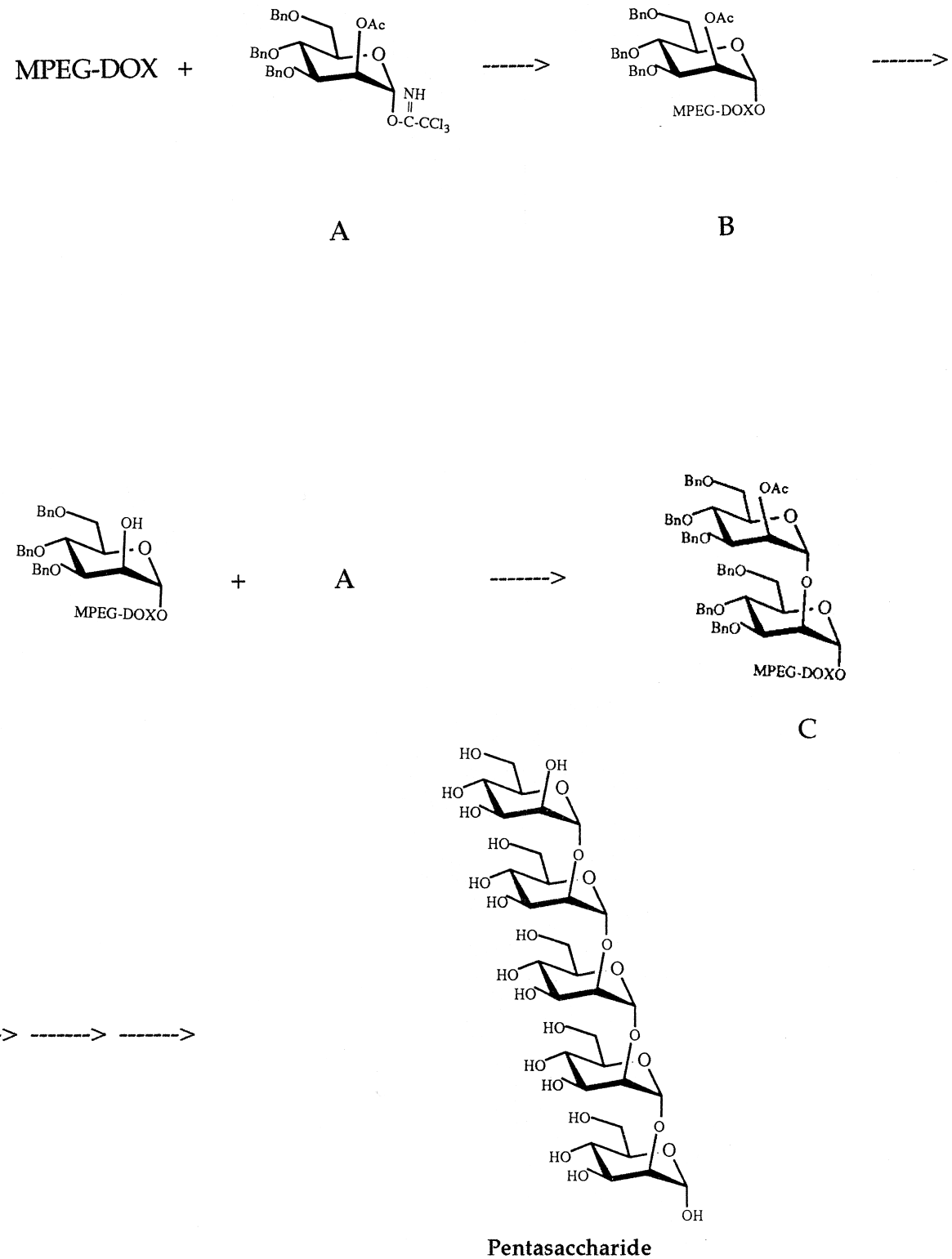
(MPEG)

$$\text{Me-O-(CH}_2\text{-CH}_2\text{)}_n\text{-CH}_2\text{-CH}_2\text{-O-CH}_2\text{-(C}_6\text{H}_4\text{)-CH}_2\text{-Cl} \rightarrow$$

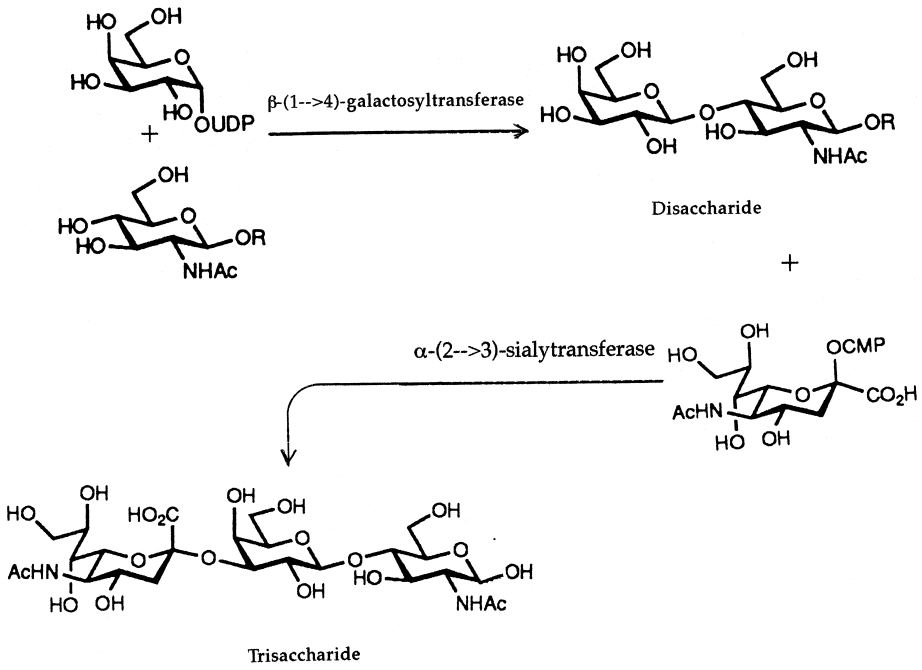
$$\text{Me-O-(CH}_2\text{-CH}_2\text{)}_n\text{-CH}_2\text{-CH}_2\text{-O-CH}_2\text{-(C}_6\text{H}_4\text{)-CH}_2\text{-OH}$$

(MPEG-DOX)

Scheme 25 illustrates the use of MPEG-DOX in the synthesis of a pentasaccharide. The first monosaccharide added (A) is a benzyl blocked trichloroacetimidate



SCHEME 25 Synthesis of a pentasaccharide using a fixed bed of polymer (MPEG) attached to a linker (DOX). The pentasaccharide is separated at the end of the synthesis by hydrogenation, which also rermoves the Bn blocking groups.



SCHEME 26 Biochemical synthesis of a trisaccharide using two enzymes, β-(1 → 4)-galactosyltransferase to form the disaccharide and an α-(2 → 3)-sialytransferase to form the desired product.

having an acetyl group where the next monomer is to be attached. Its reactive leaving group, the trichloroacetimidate, is replaced by the OH of MPEG-DOX to give a blocked monomer-MPEG-DOX (B), which is treated with base (DBU) to remove the acetyl (Ac) group and then with A to form a blocked dimer (C). Repetition of these operations will form the deacetylated blocked trimer, tetramer, and pentamer. When the desired number of monomers are linked together, the product is catalytically hydrogenated to remove the benzyl blocking groups and separate the oligosaccharide from the linker and polymer (MPEG-DOX). The pentasaccharide is separated at the end of the synthesis by hydrogenation, which also removes the Bn blocking groups (see Scheme 25).

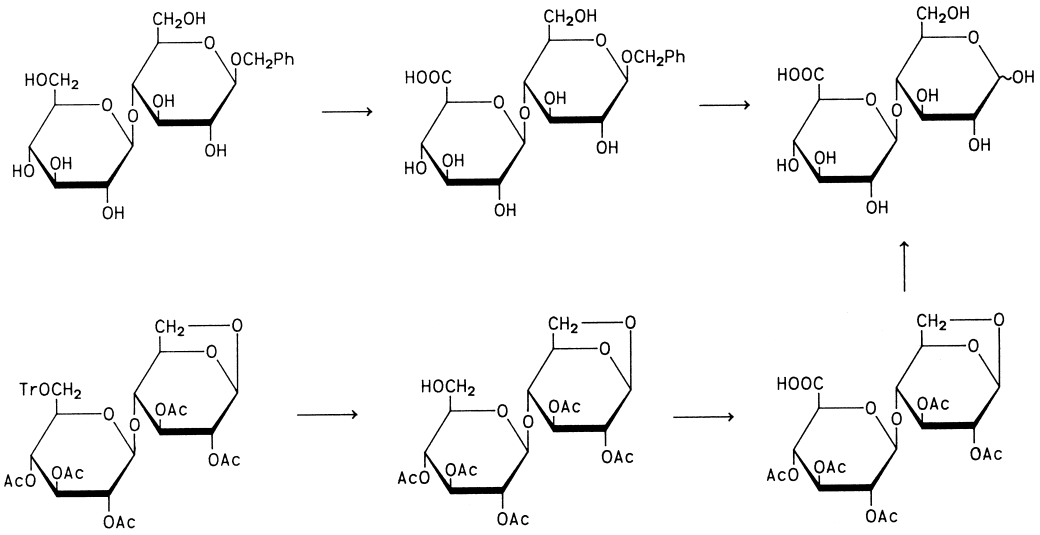
3. Biochemical Synthesis of Oligosaccharides

This method is particularly useful for the synthesis of complex oligosaccharides that exist in small amounts difficult to characterize and study. Examples of these are the cell surface oligosaccharides responsible for cell recognition. The most commonly used enzymes are glycosidases and glycosyltransferases, or a combination of both. The glycosidases are hydrolytic enzymes that can used for synthesis by carrying out the reaction in the reverse direction. Glycosyltransferases are enzymes that catalyze the synthesis of glycosides and oligosaccharides by the transfer of monosaccharides from nucleotide donors to specific substrate acceptors. The above scheme illustrates the enzymatic synthesis of a trisaccharide by using a β-(1 → 4)-*galactosyltransferase* to form a disaccharide and an α-(2 → 3)-*sialytransferase* to form the target trisaccharide (see Scheme 26).

4. Reactions

Chemical modifications of oligosaccharides usually involve the reactive primary hydroxyl group, because selective blocking and deblocking of secondary hydroxyl groups is considerably more difficult. As the DP of an oligosaccharide increases, directing a substituent in a particular moiety becomes more difficult. In a disaccharide, it is possible to direct a substituent toward the terminal (nonreducing) saccharide moiety by blocking the primary hydroxyl group of the reducing moiety with a 1,6-anhydro ring. The latter is introduced by treatment of the disaccharide glycoside (usually a phenyl glycoside) with base. Two examples of selective reactions involving the terminal ring of a disaccharide are depicted in Scheme 27.

1. The introduction of a carboxylic group by oxidation of the primary hydroxyl group of β-benzyl cellobioside with oxygen in the presence of palladium on charcoal occurs preferentially at the terminal ring. However, greater selectivity can be reached if the 1,6-anhydro derivative is first formed.



SCHEME 27 Selective reactions of the nonreducing ring of cellobiose.

2. The introduction of an azido group in the 6′ position of lactose can also be achieved by tosylating the 1,6-anhydro derivative of cellobiose, tritylating, and fully acetylating the product (Scheme 28). When the trityl group is removed by acid, the acetyl group attached to position 4 migrates, yielding a 4-hydroxy-6-*O*-acetyl derivative. Mesylating at position 4 and displacing the leaving group with azide affords the 4-azido derivative of lactose. This azido group can be reduced to an amino group or can be photolyzed to afford an aldehydic group.

Of considerable importance as surfactants are the fatty acid esters of disaccharides. An example of these is sucrose monooleate, which possesses a hydophilic–lipophilic balance of ∼15 and is ideal for use as a mild shampoo. Such esters are obtained by reacting the disaccharide in DMF with the methyl ester of the fatty acid.

IV. POLYSACCHARIDES

Polysaccharides are polymeric saccharides linked by acetal (glycosidic) linkages. The repeating units are monosaccharides or oligosaccharides, seldom larger than pentasaccharides. Although some polysaccharides are made up of fewer than a hundred sugar residues, the majority are much larger, reaching several thousand monomer units in length.

Due to their high molecular weights, most polysaccharides exhibit the characteristic properties of high polymers (low solubility and high viscosity, etc.). This is why the DP of polysaccharides can be determined by the same methods used for the molecular weight determination of synthetic polymers.

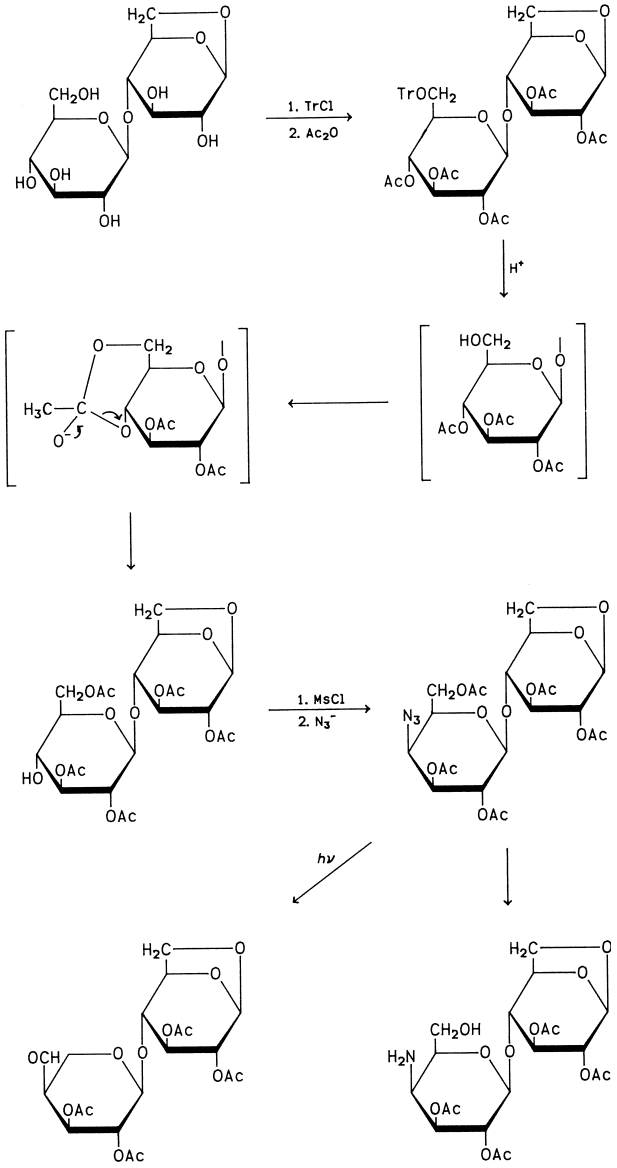
Polysaccharides are usually given names that reflect their origin, for example, cellulose, the principal component of plant cell walls. Otherwise, a systematic nomenclature can be used that suffixes the ending "an" to the name of the monomer. For example, glycan and mannan are polymers of glucose and mannose, respectively, whereas galactomannans are copolymers of galactose and mannose. Many early polysaccharide names that have other endings, such as pectin and chitin, are still in use.

Polysaccharides may exist as homologous series that have average molecular weights, rather than possess discrete molecular weights as proteins do.

A. Structure

The types of monosaccharides found in polysaccharides can easily be determined after depolymerization (by acid-catalyzed hydrolysis), by chromatographic analysis. Liquid chromatography may be used directly on the hydrolysate, whereas gas chromatography is usually carried out after sililation in order to volatilize the monosaccharides.

Because polysaccharides posses a certain naturally imposed simplification, their structure elucidation is much simpler than would appear at first glance. A further simplification in polysaccharide structure is that only a very few monosaccharides are found in natural polysaccharides. There are only four commonly occurring hexoses (glucose, mannose, fructose, and galactose), two



SCHEME 28 Conversion of a cellobiose derivative to a 4-amino-lactose.

pentoses (xylose, arabinose), two 6-deoxyhexoses (rhamnose, fucose), and two amino sugars (glucosamine and galactosamine). Four uronic acids (glucuronic, galacturonic, mannuronic, and iduronic acids) are found in several polysaccharides. Of these, D-glucuronic acid and D-galacturonic acid are the most common. The other two, D-mannuronic acid and L-iduronic acid, are found in alginic acids and in heparin, respectively.

Polysaccharides that afford on hydrolysis only one monosaccharide type are termed homoglycans, whereas those that afford two or more monosaccharide types are termed heteroglycans. Each of these can be either linear or branched.

1. Linear Polysaccharides

Linear polysaccharides are formed when the hemiacetal hydroxyl group (on C-1 of an aldose) is replaced (substituted) by a hydroxyl group from an adjoining monosaccharide unit. When this is repeated, a linear chain is formed that has at one end (called the reducing end) a reducing monosaccharide (one that has a free C-1 hydroxyl group) and at the other end (referred to as the nonreducing end) a glycosidically linked saccharide.

The glycosidic linkage in polysaccharides is usually repeated in a regular manner, with little or no randomness. However, even such seemingly regular molecules as cellulose and amylose have irregularities in their structures. Rare as these may be (sometimes the frequency is only one in a thousand), these irregularities are real. They are attributable to the fact that within the enzyme–substrate system there do occur during the course of time changes, either through abnormal action on the part of the principal chainsynthesizing enzyme or through interference of a second enzyme. The most abundant linear polysaccharide is cellulose, a glucan linked by β-D-(1 $\rightarrow$ 4) bonds, which is present in greater quantity than all other polysaccharides combined. Another important linear polysaccharide is amylose, which is a glucan linked by α-D-(1 $\rightarrow$ 4) linkages.

2. Branched Polysaccharides

When the hemiacetal hydroxyl groups of two monosaccharides are replaced by two hydroxyl groups belonging to a third sugar residue, a branching point is produced in the molecule, which then becomes a branched polysaccharide. The molecule may contain a single or numerous branch points. Examples of branched polysaccharides are amylopectin (the branched component of starch) and the gums and mucillages.

B. Polysaccharides of Industrial Importance

1. Starch

Most plants produce starch, but only a few concentrate it in amounts suitable for industrial production. These plants include cereals, such as corn, which may contain up to 80% starch, grains such as rice and tubers such as potatoes. Starch is found in the form of microscopic granules which possess characteristic shapes used to identify their source.

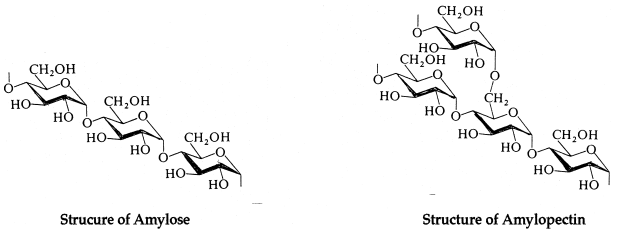


FIGURE 16 Partial structures of the two forms of starch; the linear amylose and the branched amylopectin.

The industrial preparation of starch starts by loosening the starch granules away from their matrices with cold water and letting them settle in tanks before filtration and drying. Purification of starch is achieved by repeated cycles of dispersion in cold water, sedimentation, and filtration. If desired, such treatment may be followed by fractionation to separate the linear polymer of starch, amylose from the branched polymer, amylopectin (see Fig. 16).

Amylose: Amylose is the linear form of starch; it is precipitated from a cold aqueous dispersion of starch granules with alcohols such as thymol or 1-butanol and filtered. The accompanying amylopectin is left in the supernatent solution and if needed can be precipitated. Usually the ratio of amylose to total starch ranges between 15 and 35%, however, certain plant hybrids have been developed which are capable of producing a much larger excess of one polymer over the other. Thus, amylomaize contains more than 80% of amylose; whereas waxy maize starch contains only 2% of this linear polymer. Amylose is composed of $(1 \rightarrow 4)$-linked α-D-glucopyranosyl residues attached by acetal bonds. Due to lengthy manipulation during end group assay, the DP of amylose measured chemically is invariably lower than the DP measured by physical methods, such as light scattering or ultracentrifugation. For example, the DP calculated from the amount of glucose present at the nonreducing end of the chain and estimated from the amount of 2,3,4,6-tetra-*O*-methyl-D-glucose produced upon methylation and hydrolysis of amylose ranges between 200 and 300. This is much lower than the DP obtained by light scattering or ultracentrifugation (about 6000). The difference has been attributed to degradation during the chemical treatment. In neutral solution amylose exists as a random coil, but other conformers can be produced by retrogradation, i.e., separating the different insoluble fractions of amylose that deposit on standing. In the presence of complexing agents amylose assumes a more organized conformation; namely that of a right-handed helix made up of 150 to 1000 turns and containing six α-D-glucopyranose rings per turn. The dimensions of the amylose helix are such that it can accommodate an iodine atom, or a lipid molecule of appropriate size. The light absorption of iodine-amylose complexes is very similar to that of iodine in nonoxygenated solvents, such as hydrocarbons. This observation has been attributed to the lack of interaction between the equatorial oxygen atoms of the six α-D-glucopyranose rings forming a turn of the helix and an atom at its center. The latter is in a lipophilic surrounding (of the axial C-H groups), very similar to that of iodine in a hydrocarbon solution. The lipophilic center of the helix would also make it difficult to remove a lipid molecule located in its center, which explains why the ^{13}C-NMR spectrum of pure amylose invariably shows lipid absorptions.

Amylopectin: Amylopectin is a highly branched α-D-glucose polymer that possesses a much larger molecular weight than amylose. It contains a number of linear chains of disubstituted $(1 \rightarrow 4)$-linked α-D-glucopyranosyl residues similar to those found in amylose, in addition it contains some $(1 \rightarrow 6)$ linked branches originating from 1,4,6-tri-*O*-substituted α-D-glucopyranosyl residues. Light scattering data gives amylopectin a DP value in the order of 10^5 or higher if the starch from which it was obtained had matured for a long period in the plant. The accepted shape of the amylopectin macromolecules is that of a three-dimensional tree or bush (shown two dimensionally in Fig. 17). This conformation is based on data obtained from methylation experiments and enzymatic methodologies. For example, methylation of amylopectin yields a smaller amount amount of 2,3,6-tri-*O*-methyl-D-glucose than amylose. Because 2,3,6-tri-*O*-methyl-D-glucose is produced from linear portions of $(1 \rightarrow 4)$ linked α-D-glucopyranose polymers and because it is formed in a smaller ratio from amylopectin than from amylose, one can conclude that the chains of amylopectin are considerably shorter than those of amylose. The high degree of branching of amylopectin can also be deduced from the relatively large amount (4%) of 2,3,4,6-tetra-*O*-methyl-glucose, produced from the nonreducing ends of the amylopectin molecule, as compared to the amount produced from amylose (<0.4%). In addition, amylopectin yields a unique product not usually found when amylose is methylated and hydrolyzed, namely of 2,3-di-*O*-methyl-D-glucose, produced from the points of branching.

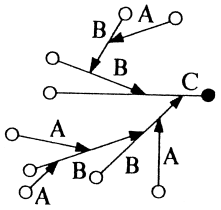


FIGURE 17 A two-dimensional depiction of the bush model of amylopectin. The reducing end of the molecule is represented by a solid black circle; the other circles represent the nonreducing ends.

a. Starch-based industries. Dry milling is used to produce flours such as corn flour and meals such as oatmeal, whereas wet milling, is used as feedstocks in the production of sugar syrups, used in the manufacture of foods, soft drinks, and beer.

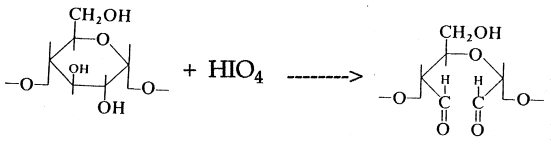
Sweeteners From Starch: The most widely used natural sweetener is the disaccharide sucrose, which is available in crystalline form as well as in the form of syrups under various names such as molasses and invert sugar. Up until recently the only serious competitor of sucrose was the cheaper glucose syrup produced since the early nineteenth century by the acid hydrolysis of starch and more recently by its enzymatic hydrolysis with α-amylases and glucoamylases. Because glucose syrup is significantly less sweet than sucrose, greater quantities of this syrup must be used to yield a product of comparable sweetness to sucrose. This is a serious disadvantage, duly noted by the calorie-conscious public and by the food industry, which took steps to remedy the situation. The glucose syrup was partly converted to fructose by the enzyme glucose isomerase obtained from *Bacillus coagulans*. The resulting product, designated fructose syrup, contains 42% fructose and possesses nearly the same sweetness as sucrose. Recently, this sweetener was further improved upon; it was converted into two products that are sweeter than sucrose. They are very high fructose syrups that contain about 90% fructose and crystalline fructose. Both are obtained from fructose syrup by purification and separation on ion exchange columns. Fructose syrup and high fructose syrup have now replaced glucose syrup in many soft drinks, jams, confectionery, and conserves and crystalline fructose is slowly replacing sucrose. It is estimated that half the sweeteners consumed in the United States are now produced by the hydrolysis of starch in continuous flow biocatalyst reactors that efficiently reuse the enzymes.

Starch Fermentation Industries: Beer is manufactured from barely whereas sake is obtained from rice. Germinated barely is used in breweries to produce the enzyme β-amylase needed for the breakdown of starch, whereas a mold, *Asparagillus oryzea*, is used to produce the α-amylase needed to hydrolyze rice starch for the manufacture of sake. The wort in both cases is subjected to fermentation with the yeast, *Saccharomyces serevisea*. As in many modern food-related industries, continuous flow reactors and immobilized enzymes have been used in the manufacture of beer, and more recently of wine, and champagne. This has resulted in a considerable reduction in cost and residence time.

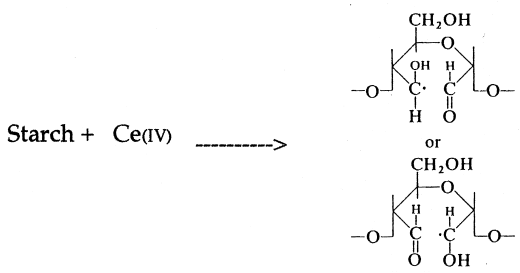
Starch and starch-rich wastes from the food industry have been proposed as feedstocks for the fuel industry to produce gasohol, and the food industry to produce single-cell proteins for animal consumption.

b. Chemically modified starch. The properties of starch are often ameliorated by chemical modification. This is why today the majority of the starch produced for industry is chemically modified. Examples of such modified starches shown below.

Oxidized Starch: Starch is oxidized to remove some of its numerous OH groups that are responsible for hydrogen bonding and for gel forming. The resulting starch solutions after oxidation form weaker gels and are less susceptible to autodecomposition or retrogradation. The oxidations are usually carried out in basic media using oxidants such as sodium hypochlorite or better sodium periodate. Oxidation with the latter is usually followed by electrolysis to regenerate periodic acid from the iodic acid formed during the process. The product formed by periodate oxidation is a 2,3-dialdehyde formed by cleavage of the 2,3- bonds of starch (see Scheme 29). This can be made to react with



Dialdehyde + Cellulose --------> Graft Polymer



Free Radical + Acrilonitrile ---------> Graft Polymer

SCHEME 29 Oxidation of starch with periodic acid gives a dialdehyde, which can be grafted to natural polymers such as cellulose. Oxidation of starch with ceric ions gives a free radical which can be grafted to acrilonitrile to give an absorbant graft polymer.

natural or synthetic polymers having OH groups to form cross-linked polymers. Cross linking with cellulose affords products possessing enhanced wet strength that are used in the paper industry to make paper towels. Cross linkage with cotton affords shrink- and crease-resistant fabrics for the textile industry. It is also possible to react the free radicals produced by oxidation of starch with ceric ions with monomers to form graft polymers (see Scheme 29). For example, copolymerization of oxidized starch with polyacrylonitrile yields excellent absorbents used in the manufacture of disposable diapers capable of absorbing 5000 times their weight in water.

Cross-Linked Starches: Cross-linked starches are obtained by forming bridges between the hydroxyl groups of the starch molecule and the hydroxyl groups of another molecule. The resulting compounds are extensively used by the food industry as thickeners or to impart added mechanical stability to prevent the food from loosing its shape during cooking. In theory any molecule possessing two groups capable of reacting with hydroxyl groups can be used as a bridge between starch molecules. An example of such bridge-forming molecules is epichlorhydrin, which forms distarch glycerol depicted below.

$$\text{Starch-OH} + \underset{\diagdown\,\text{O}\,\diagup}{\text{CH}_2\text{-CH-CH}_2\text{-Cl}}$$

Epicholorohydrin

$$\longrightarrow \text{Starch-O-CH}_2\text{-}\underset{\displaystyle\text{OH}}{\underset{|}{\text{CH}}}\text{-CH}_2\text{-O-Starch}$$

Distarch glycerol

Starch Ethers: Starch ethers are produced by reacting starch with ethylene or propylene oxide to the desired DS. For example, starch hydroxypropyl alcohol is produced by reacting starch with propylene oxide. The nucleophilic substitution is carried out in basic media and the resulting starch ether, depicted below, is widely used in paper, textile, and food industries.

$$\text{Starch-OH} + \underset{\diagdown\,\text{O}\,\diagup}{\text{CH}_2\text{-CH-CH}_3} \longrightarrow \text{Starch-O-}\underset{\displaystyle\text{OH}}{\underset{|}{\text{CH}_2}}\text{-CH-CH}_3$$

Propylene oxide Starch ether

2. Glycogen

Glycogen is a homopolysaccharide found in the liver and muscles of animals, where it is used to store energy. Chemically, glycogen is related to starch and closely resembles amylopectin. It is composed of linear chains of disubstituted (1→4)-linked α-D-glucopyranosyl residues attached to (1→6) linked branches originating from 1,4,6-tri-*O*-substituted α-D-glucopyranosyl residues. The difference between glycogen and starch is in the source (starch is found in plants) and the shape of its aggregates and their size. Instead of starch granules, glycogen is found in the form of spheres called β-partcles, having a DP of 10^5 and aggregates of spheres called α-particles having a DP of 10^7. The shape of the glycogen macromolecules resembles a bush or a tree (see Fig. 18). However, its agregates are much more symmetric than those of amylopectin (see Fig. 17). This is why the aggregates of starch are called granules and those of glycogen, spheres.

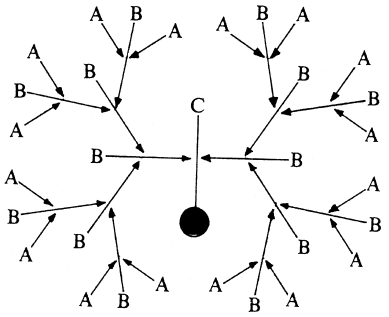


FIGURE 18 Bush structure of glycogen. Note that the molecule is more symmetric than that of amylopectin.

3. Dextrans and Glucans

Dextrans are α-D-glucopuranose polymers linked for the major part through (1 → 6) acetal bonds as well as by some (1 → 3) and (1 → 4) linkages. Dextrans are highly branched polymers having a DP in excess of 10^6. Although dextrans are obtained from a multitude of bacteria, most of the industrially produced products are obtained from *Leuconostoc* mesenteroides and *Leuconostoc dextranicum*.

D-*glucans* are obtained from fungi and may either be α-D-plucopyranose or β-D-glucopyranose polymers; they contain mainly (1 → 3) and (1 → 4) glycosidic linkages and rarely (1 → 6) bonds.

4. Cellulose

Cellulose is the most abundant organic compound on earth; its amount exceeds by far that of any other organic polymer or monomer. Cellulose is found in a nearly pure form in cotton and flax and in much larger amounts, but less pure form in wood, from which it can be produced by dissolving its contaminants, namely hemicellulose and lignin, with alkali. Cellulose-based industries include the pulp and rayon industries where it is produced and purified, respectively, and the paper and the textile industries, where it is processed. Like amylose, cellulose is a linear polymer made up of (1 → 4)-linked D-glucopyranosyl residues, which differ from those of

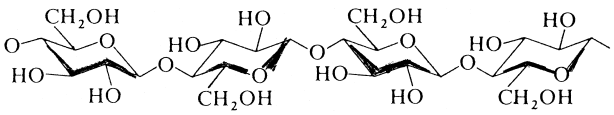


FIGURE 19 Structure of cellulose.

amylose in their anomeric configuration; β for cellulose and α for amylose. The DP of cellose, determined by physical methods is around 10^4, much higher than the value deduced chemically by end group analysis, which is only 200. The difference is attributed to its degradation during chemical manipulation. A similar difference was observed with amylose. Upon methylation and hydrolysis, cellulose affords 2,3,6-tri-*O*-methyl glucopyranose in high yield (>90%). The presence of (1 → 4)-linkages can be confirmed by partial hydrolysis to the disaccharide, cellobiose, whose β-D-configuration was confirmed by enzymatic hydrolysis (for a partial structure of cellulose see Fig. 19). The actual shape of the cellulose aggregates was studied by X-ray diffraction, which revealed that there exist two types of cellulose: cellulose I and cellulose II. Cellulose I is formed naturally and is composed of highly ordered microfibrils that are held together by hydrogen bonds and are further associated to form fibers. Bundles of cellulose I all have their reducing ends in the same side of the macromolecules and the nonreducing end in the opposite side. A different type of bundle occurs in cellulose II, found in regenerated cellulose such as viscose rayon. The aggregates of cellulose II have their reducing ends arranged in an antiparallel manner, i.e., their reducing ends occur at both ends of the bundle. Cellulose II types can be further divided into III_I and III_{II} and IV_I and IV_{II}, according to the conformation of their unit cells; the I family forms bent chains, whereas the II families adopt bent-twisted conformations.

a. Chemically modified cellulose. The most important derivatives of cellulose are the esters, such as cellulose acetate and nitrate, the ethers, such as carboxymethyl, hydroxyethyl, and hydroxypropyl cellulose, and the rayons, of which the most important are the acetate rayon, which is discussed next and viscose rayon obtained from cellulose xanthate.

Cellulose Acetate: Cellulose acetate, also known as acetate rayon is obtained from wood pulp in three steps: (1) pretreatment (soaking with acetic acid and sulfuric acid); (2) acetylation with a mixture of acetic anhydride and acetic acid, to form the 2,3,6-tri-*O*-acetyl derivative; and (3) partial hydrolysis to afford a product having a DS of about 2.6, which forms the best fibers and films.

Cellulose Nitrate: Cellulose can be esterified with a mixture of nitric acid and some sulfuric acid added as catalyst. The amount of water controls the degree of substitution, which in turn determines the appropriate usage for the product; higher DS products are used as explosives and lower ones as cements and lacquers.

Carboxymethyl Cellulose: This ether of cellulose is obtained by a modified Williamson synthesis whereby cellulose is first treated with sodium hydroxide and the alcoholate formed is reacted with the halide, namely sodium monochloroacetate. Although a DS of 3 is attainable (positions 3,4, and 6 can all be etherified), a DS of only 0.4 is all that is needed to achieve the desired viscosity and solubility needed in the commercial products. Carboxmethyl cellulose is widely used in the form of alkali metal salts as surfactant in detergents, in oil drilling, mining, and in the food, paper, and textile industries.

Hydroxyethyl Cellulose: Ether alcohols such as hydroxyethyl cellulose are used industrially to solubilize cellulose and to facilitate its handling. Hydroxyethyl cellulose is prepared by swelling the cellulose with sodium hydroxide and reacting it with ethylene glycol. As in the previous example, a low DS of 0.3 is sufficient to solubilize the product in alkali.

$$\text{Cell-OH} + \underset{\diagdown\,\text{O}\,\diagup}{\text{CH}_2\text{-CH}_2} \longrightarrow \text{Cell-O-CH}_2\text{-CH}_2\text{-OH}$$

Cellulose — Hydroxyethyl Cellulose

Hydroxypropyl cellulose is prepared in the same manner except that propylene glycol is used in place of ethylene glycol.

Cellulose Xanthate: Cellulose xanthate or viscose is an intermediate in the production of viscose rayon, used in the manufacture of textile fibers and cellophane wrapping. The manufacture of viscose starts by swelling the cellulose with sodium hydroxide. The mixture is then treated with carbon disulfide to form the sodium xanthate of cellulose, which is known as viscose. It was given this name because its viscosity increases with time. When the desired consistency is reached, the mass is extruded into an acid bath to form fibers or sheets depending on the extrusion orifice shape (circular holes or slits). The acid bath decomposes the xanthate groups and regenerates cellulose now called viscose rayon to distinguish it from other rayons.

$$\text{Cell-OH} \longrightarrow \text{Cell-ONa} + \text{CS}_2 \longrightarrow \text{Cell-OCSSNa}$$

Cellulose — Viscose

$$\longrightarrow \text{Cell-OH}$$

Viscose rayon

5. Mannans and Galactomannans

Mannans are highly insoluble polymers of β-D-mannopyranose, found in the kernels of many palms, for example, ivory nut. This hard material is composed of nearly

pure mannan, and was, until the advent of plastics, used for making buttons. Many palm kernels contain less pure polymers of mannose, often containing galactose and referred to as galactomannans.

6. Chitin and Chitosans

Chitin and Chitosans are polymers of β-D-glucosamine; the main difference between them is in DP, the first polymer has a DP around 10^5 whereas the second has only 10^4. As a result chitosan is much more soluble than chitin and is therefore easier to handle and to shape; it is used in photographic films, for immobilization of enzymes, and its phosphate esters as fire retardants.

7. Hemicelluloses

Hemicellulose is the most abundant naturally occurring organic compound after cellulose. It differs from cellulose in that it dissolves in alkali to form a dark brown solution known in the paper industry as "black liquor." Some hemicelluloses are linear while others are highly branched. For example, *xylans*, the most common component of hemicellulose, are mostly linear polymers made up of $(1 \rightarrow 4)$ linked β-D-xylopyranosyl residues, which may occasionally branch at the 3-position and *arabino*-D-*galactans* which are highly branched polymers.

8. Pectin

Pectin is produced in large quantities as by-products of the apple and orange juice productions. The pulp of apples and the peels of orange left over after pressing contain about 15 and 30% pectin, respectively. The pectin extracted from these agricultural wastes is used in the manufacture of jams and jellies. It is also used as a stabilizer of fruit drinks and pharmaceutical suspensions because of its great stability at low pH.

V. OLIGONUCLEOTIDES AND POLYNUCLEOTIDES

Oligonucleotides are DNA or RNA segments of low molecular weight; they are composed of nucleotide monomers (see Fig. 20) linked by phosphoric ester bridges spanning C-5′ of one unit and C-3′ of the other. Synthetic oligonucleotides have been widely used in the study of DNA, in protein biosynthesis, and in induced mutagenesis. They were instrumental in deciphering the genetic code, and they have since been used to introduce mismatches in specific sites of DNA strands to induce specific changes in enzymes.

More recently, antisense oligonucleotides have been used in the fight against AIDS and cancer. Anitisense oligonucleotides were synthesized that complement portions of *m*RNA which translate to the protein cover of the HIV virus or to a cancer specific oncogene. In the cell, the oligonucleotides complex with the targeted RNA and arrest the synthesis vital proteins (see Fig. 21).

A. Synthesis of Oligonucleotides

Both oligodeoxyribonucleotides (DNA-type oligomers) and oligoribonucleotides (RNA-type oligomers) have been synthesized, but the first have attracted most of the

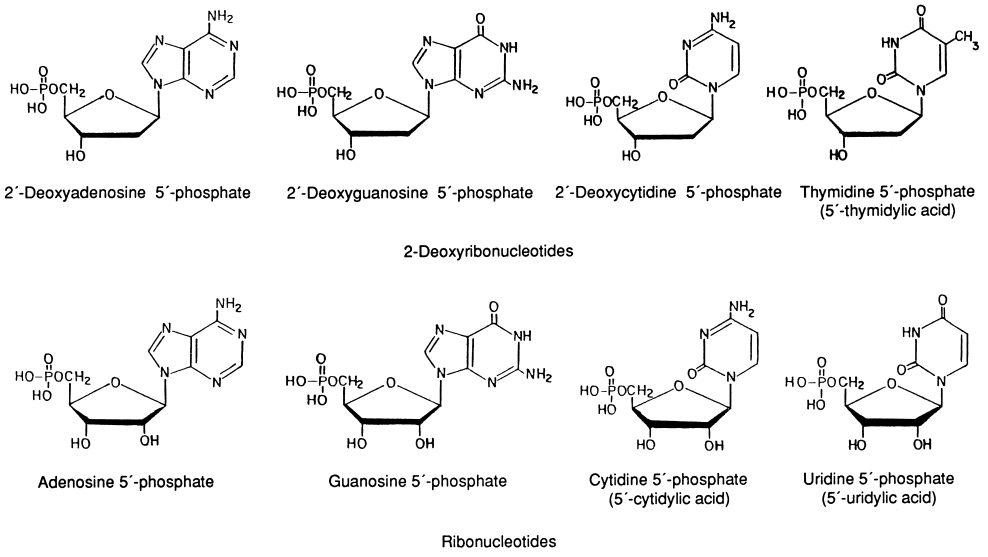


FIGURE 20 The four 2-deoxyribonucleotides found in DNA and the four ribonucleotides found in RNA.

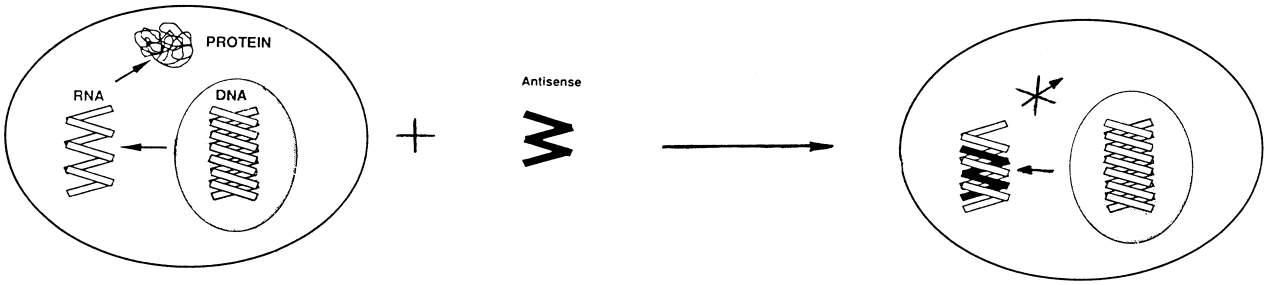
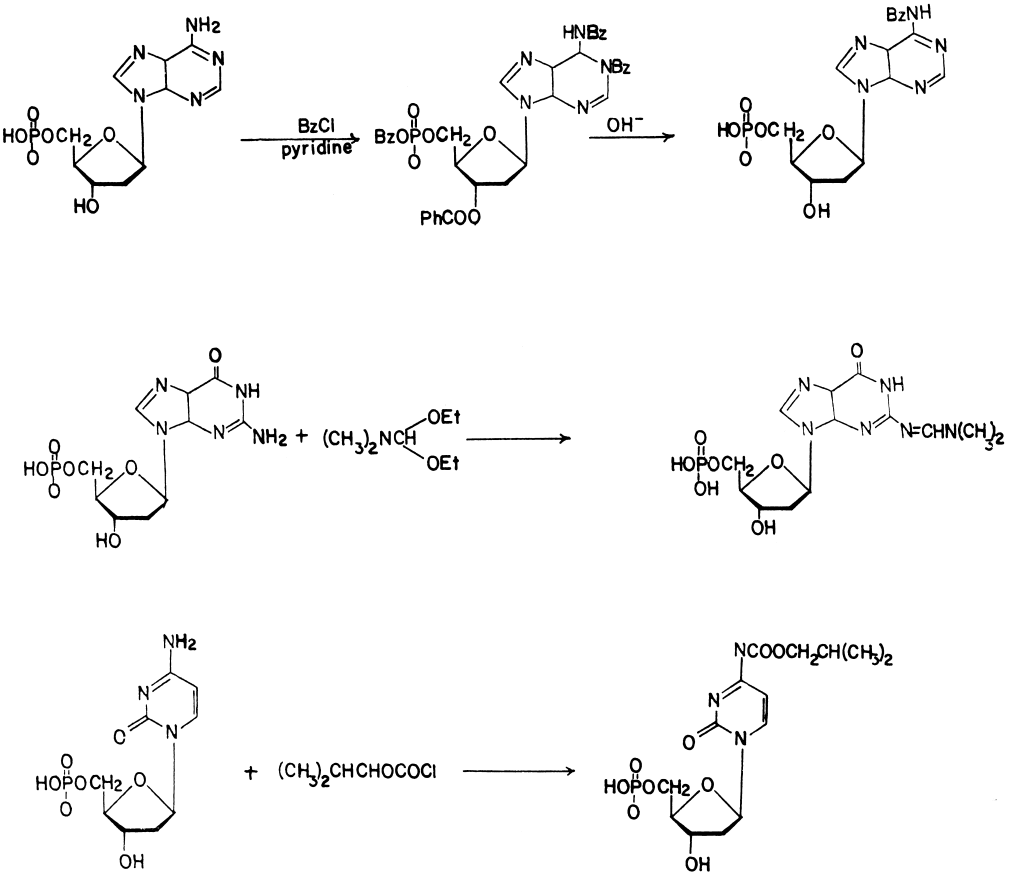


FIGURE 21 Inhibition of gene translation with antisenese oligonucleotides. An oligonucleotide designed to complexes with a segment of *m*RNA is used to arrest the synthesis of a vital protein.

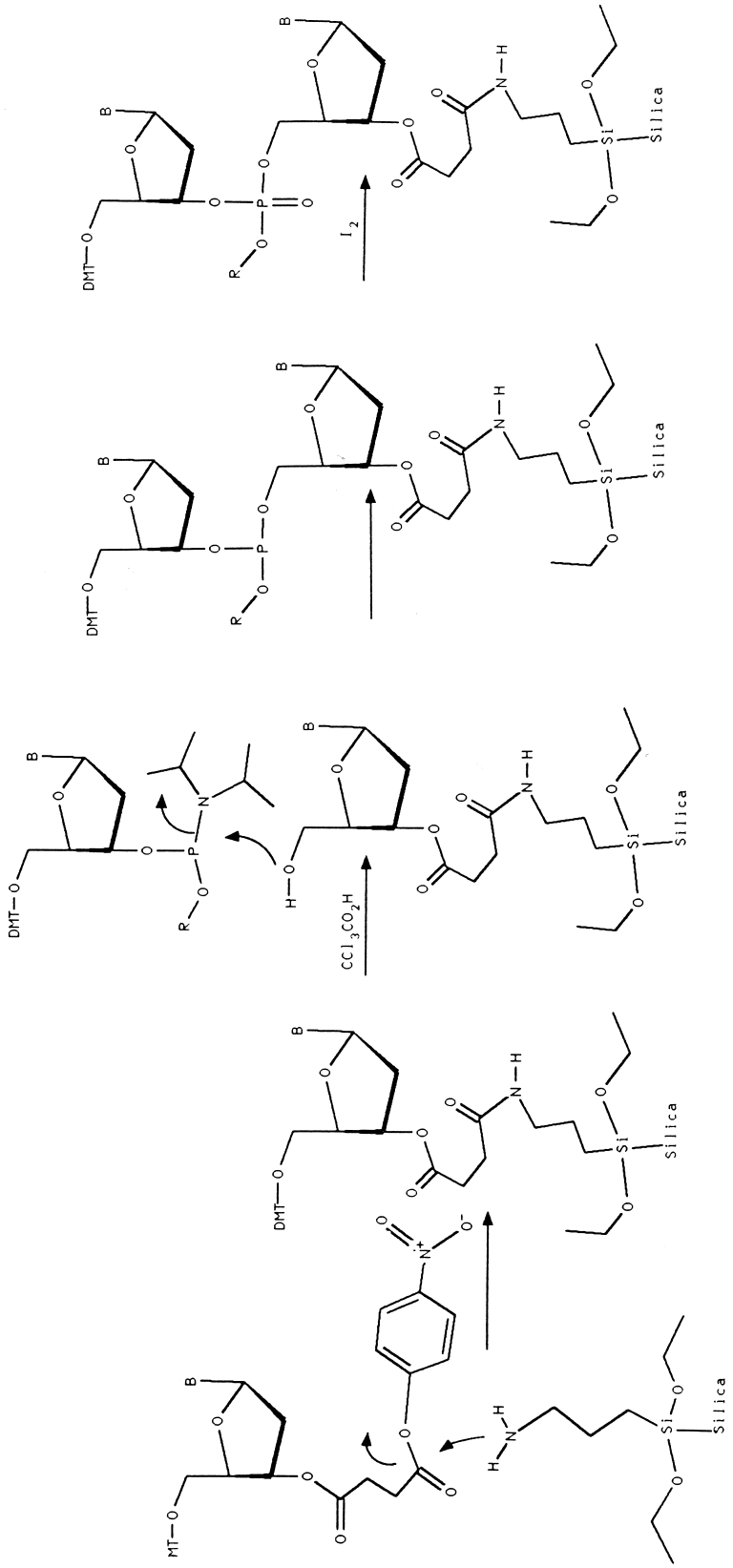
attention of chemists because of their biological importance and their relative ease of preparation. These characteristics become evident when the structures of the two types of nucleotide monomers are examined. Thus, the DNA monomers have only one free hydroxyl group (on C-3′), which, when esterified by the phosphate group of another monomer, will form a phosphate bridge in the desired position (linking C-3′ of one nucleotide to C-5′ of the other), whereas ribonucleotides have two free hydroxyl groups (one on C-2′ and one on C-3′), which necessitates that the 2′-OH group be protected to prevent the formation of the undesired $2' \rightarrow 5'$ phosphate bridge.

B. Block Synthesis of Oligonucleotides

An oligonucleotide chain may be terminally phosphated (esterified) on C-5′ or C-3′. Oligomers of the first type are prepared from natural nucleotides by treating an acetylated



SCHEME 30 Some reactions of ribonucleotides.



SCHEME 31 Formation of oligonucleotides in automatic synthesizers using fixed beds.

nucleotide (one having a phosphate group on C-5′ and an acetyl group on O-3′) with a nucleotide or an oligonucleotide having a blocked phosphate group on C-5′ and a free hydroxyl group in position 3′. The second type, those terminally phosphated at C-3′, are usually prepared from nucleosides phosphated at C-3′ (which are not available commercially). The monomers are acetylated on O-5′ and treated with other monomers having free hydroxyl groups on C-5′ and a blocked phosphate group on C-3′ to obtain the desired oligonucleotide.

In both cases, it is necessary to block the phosphate groups to prevent the formation of anhydrides (pyrophosphates). This is done by treating the nucleotide with 3-hydroxypropanonitrile to obtain an ester removable with alkali (β elimination). Other phosphate-blocking groups include $PhCH_2{-}O{-}(CH_2)_2{-}NH_2$, which forms an amide hydrolyzable by acids.

The coupling of the two nucleotide moieties involves formation of a phosphoric diester from a monoester (or a triester from a diester). The reaction is catalyzed by such condensing agents as dicyclohexylcarbodiimide (DCC), a reagent extensively used in peptide synthesis, or arylsulfonyl chlorides (for example, 2,4,6-trimethylphenyl- or 2,4,6-triisopropylphenylsulfonyl chloride).

All nucleotides possessing primary amino groups (three of the four bases present in DNA and RNA nucleotides), namely adenylic, guanylic, and cytidylic acid, must have their amino groups protected to prevent the formation of amide ester bridges (instead of diester phosphate bridges). The primary amino group of adenylic acid is usually protected by a benzoyl (or *p*-methoxybenzoyl) group. This is achieved by treating the nucleotide with an excess of benzoyl chloride and removing the undesired benzoyl groups with alkali. Guanylic acid is usually protected by reaction with *N*,*N*-dimethylformamide diethyl acetal to give a readily removable Schiff base. Finally, cytidylic acid is protected by formation of a carbamate (see Scheme 30).

C. Formation on Automated Oligonucleotide Synthesizers

The phosphite triester approach is commonly used with automatic synthesizers. The phosphite ester formed is oxidized at the end of each sequence to a phosphate ester. Typically, a deoxynucleoside 3′-phosphoramidite is added to a growing DNA segment that is linked to a silica support. The silica is first reacted with a 3-aminopropyl triethoxysilane linker and then with an appropriate tritylated nucleoside, such as a 5′-dimethoxytrityl-2′-deoxynucleoside having a 3′-*p*-nitrophenylsuccinoyl group. The addition of successive deoxynucleotides to this support-bound deoxynucleoside is carried out by (a) removal of the trityl (DMT) protecting group with trichloroacetic acid; (b) condensing with a tritylated deoxynucleoside 3′-phosphoramidite, to yield a deoxydinucleoside phosphite triester; (c) oxidizing of the phosphite triester to a phosphate triester with iodine (see Scheme 31). Repetitions of this sequence have been used to synthesize DNA segments containing up to 200 deoxynucleotides. Once a synthesis is complete, the oligodeoxynucleotide is removed from the column, freed of protecting groups, and isolated by gel electrophoresis and purified by HPLC.

SEE ALSO THE FOLLOWING ARTICLES

BIOPOLYMERS • PHYSICAL ORGANIC CHEMISTRY • PULP AND PAPER

BIBLIOGRAPHY

Aspinall, G. O., ed. (1982). "The Polysaccharides," Academic Press, New York.
Boons G. J., ed. (1998). "Carbohydrate Chemistry," Academic and Professional.
Ginsberg, V., and Robbins, P. W. (1991). "Biology of Carbohydrates," Jai Press, Tokyo.
Robert M. Giuliano, R. M., ed. (1992). "Cycloaddition Reactions in Carbohydrate Chemistry," ACS Symposium Series, Washington D.C.
Gyorgydeak, Z., and Pelyvas, Z. F. (1998). "Monosaccharide Sugars," Academic Press, New York.
Hanessian, S. (1983). "Total Synthesis of Natural Products," Pergamon, Oxford,England.
Horton, D., ed. (1995). "Advances in Carbohydrate Chemistry and Biochemistry," Vol. 50, Academic Press, New York.
Horton, D., ed. (1996). "Advances in Carbohydrate Chemistry and Biochemistry," Vol. 51, Academic Press, New York.
Horton, D., ed. (1997). "Advances in Carbohydrate Chemistry and Biochemistry," Vol. 52, Academic Press, New York.
Horton, D., ed. (1998). "Advances in Carbohydrate Chemistry and Biochemistry," Vol. 53, Academic Press, New York.
Horton, D., ed. (1999). "Advances in Carbohydrate Chemistry and Biochemistry," Vol. 54, Academic Press, New York.
Kennedy, J. F., ed. (1988). "Carbohydrate Chemistry," Clarendon Press, Oxford, England.
Ogura, H., Hasegawa, A., and Suami, T. (1992). "Carbohydrate Synthetic Methods," Kondasha, Tokyo.
Postema, M. H. D. (1995). "C–Glycosides Syntheses," CRC Press, Cleveland.

Catalysis, Homogeneous

Piet W.N.M. van Leeuwen
University of Amsterdam

GLOSSARY

Coordination compound Molecule or ion comprising a metal atom or ion surrounded by ions or molecules called ligands.
Coordination number Number of ligand atoms bonded to the metal atom or ion in a coordination compound.
Electrophile Refers to the reactivity of an atom, ion, or molecule; an electron-deficient species ready to accept an electron pair from another species thus forming a chemical bond.
Enantiospecificity Preference for the formation of a single enantiomer of a compound.
Metal complex *See* Coordination compound.
Migratory insertion Insertion of an unsaturated molecule such as ethene or CO (Y) into a metal–X bond, whereby X migrates to one of the atoms of the unsaturated molecule coordinated to the metal giving M–Y–X.
Nucleophile Refers to the reactivity of an atom, ion, or molecule; an electron-rich species ready to donate an electron pair to another species thus forming a chemical bond.
Oligomerization Coupling of several molecules of the same kind forming a larger unit (oligomer).
Oxidative addition Oxidation of a metal compound accompanied by an increase of the coordination number.
Polymerization Coupling of many molecules (monomers) forming a polymer (macromolecule).
Reductive elimination (Reverse of oxidative addition.) Loss of one or two ligands from a metal complex by reduction of the metal.

THE PHENOMENON CATALYSIS has become well known owing to the use of catalysts in automobile exhausts, which ensure the complete conversion of the last traces of fuel and decompose the toxic nitrogen oxides and carbon monoxide to nitrogen and carbon dioxide. Catalysts are substances that enhance chemical reactions that are not themselves consumed in the reactions. Exhaust catalysts are heterogeneous, as the catalyst is a solid material while the reactants are in the gaseous phase. In this article we are concerned with homogeneous catalysts, that is to say that catalyst and reactant are molecularly dispersed in the same phase, often the liquid phase. Homogeneous catalysts for the gas phase are also known and a well-known example is the chlorine atom catalyzing the decomposition of ozone in our stratosphere. Here we will focus on metal catalysts in solution that have found widespread utilization in chemical industry for the manufacture of bulk chemicals, fine chemicals, and pharmaceuticals and in organic synthesis in the laboratory. The catalysts presented

are important for a variety of reactions including hydrogenation, isomerization of alkenes, oligomerization, polymerization, carbonylation, hydroformylation, hydrocyanation, metathesis, polyester formation, etc. Many catalysts display high rates and selectivities for conversions that would have taken many steps in the absence of the catalyst or would not have been possible at all. Many of the catalytic reactions show a high atom economy, i.e., a high percentage of the atoms of the starting materials ending up in the final product as by-product formation is minimized.

I. DEFINITION AND SCOPE OF HOMOGENEOUS CATALYSIS

A. Definition

Berzelius coined the term catalysis in 1836 when he had noticed changes in substances when they were brought in contact with small amounts of certain species called "ferments." Many years later in 1895 Ostwald came up with the definition that: *A catalyst is a substance that changes the rate of a chemical reaction without itself appearing into the products*. This means that according to Ostwald a catalyst can also slow down a reaction. The definition used today reads as follows: *A catalyst is a substance that increases the rate at which a chemical reaction approaches equilibrium without becoming itself permanently involved*. The "catalyst" may be added to the reactants in a different form, the catalyst precursor, which has to be brought into an active form ("activated"). During the catalytic cycle the catalyst may be present in several intermediate forms when we look more closely at the molecular level. An active catalyst will pass a number of times through this cycle of states; in this sense the catalyst remains unaltered. The number of times that a catalyst goes through this cycle is the turnover number. The turnover number (t.o.n.) is the total number of substrate molecules that a catalyst converts into product molecules. The turnover frequency (t.o.f.) is the turnover number in a certain period of time. Substrates are present in larger amounts than the catalyst; when we report on catalytic reactions the ratio of substrate to catalyst is an important figure. An inhibitor is a substance that retards a reaction. An inhibitor is also present in "catalytic" or substoichiometric amounts. In a radical chain reaction an inhibitor may be a radical scavenger that interrupts the chain. In a metal-catalyzed reaction an inhibitor could be a substance that adsorbs onto the metal making it less active or blocking the site for substrate coordination. We also talk about a poison, a substance that stops the catalytic reaction. We will often see the word co-catalyst, a substance that forms part of the catalyst or that plays another role somewhere in the catalytic cycle.

Kinetics is an important part of catalysis; after all, catalysis is concerned with accelerating reactions. Comparison of catalysts and comparison of catalyzed and noncatalyzed reactions is not a straightforward task. Suppose we have a bimolecular reaction of species A and B with a rate of product formation:

$$d[\mathrm{P}]/dt = k_1[\mathrm{A}][\mathrm{B}]$$

We don't know what the rate equation for the catalyzed reaction might look like, but it is reasonable that at least the catalyst concentration will occur in it, e.g.:

$$d[\mathrm{P}]/dt = k_2[\mathrm{Cat}][\mathrm{A}][\mathrm{B}]$$

Hence the dimension of the reaction is different, even in the simplest case, and hence a comparison of the two rate constants has little meaning. If both reactions occur simultaneously in a system we may determine what part of the product is made via the catalytic route and what part isn't. In enzyme catalysis and enzyme mimics one often compares the k_1 of the uncatalyzed reaction with k_2 of the catalyzed reaction; if the mechanisms of the two reactions are the same this may be a useful comparison. In practice the rate equation may take a much more complicated form than the ones shown above. The rate equation tells us something about the mechanism of the reaction.

Before we turn to "mechanisms" let us repeat how a catalyst works. We can reflux carboxylic acids and alcohols and nothing happens until we add traces of mineral acid that catalyzes esterification. We can store ethene in cylinders for ages (until the cylinders have rusted away) without the formation of polyethylene, although the formation of the latter is exothermic by more than 80 kJ/mol. We can heat methanol and carbon monoxide at 250°C and 600 bar without acetic acid being formed. After we have added the catalyst the desired products are obtained at a high rate.

A catalyst lowers the barrier of activation of a reaction, i.e., it lowers the activation energy. When protons or Lewis acids are the catalysts this description seems fairly accurate. Take for instance a Diels-Alder reaction catalyzed by a Lewis acid (Fig. 1):

The catalyst makes the dienophile more electrophilic. It lowers the energy level of the LUMO and the interaction between the LUMO of the dienophile and the HOMO

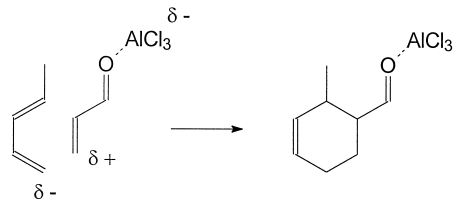


FIGURE 1 Lewis acid catalyzed Diels-Alder reaction.

of the diene increases. As a result the reaction becomes faster than the uncatalyzed one. Accidentally it becomes also more regioselective. In this instance the catalyzed reaction is very similar to the uncatalyzed reaction. Often this is not the case. In particular when the reagents are very unreactive toward one another (ethene versus ethene, methanol versus CO, see above) the simple picture of a catalyst that lowers the reaction barrier is beyond the truth.

The following description is more general. A catalyst provides a more attractive reaction pathway for the reagents in order to arrive at the products. This new pathway may involve many steps and may be rather complicated. Summarizing, *a catalyst provides a new reaction pathway* with a low barrier of activation, which may involve many intermediates and many steps. The sequence of steps is called the mechanism of the reaction. Mechanism also refers to the more detailed description of a reaction at the molecular bonding level. During the process, the catalytic cycle, the catalyst participates in many "complexes" all of which might be named "the" catalyst. It cycles continuously from one species to another. In this sense the catalyst itself remains unchanged during the catalytic conversion (Ostwald, first paragraph). The catalyst may spend most of its time in the form of a certain species, which is often the species undergoing the slowest reaction of all in the catalytic cycle. This species may be observed by *in situ* spectroscopy and it is referred to as the *resting state* of the catalyst.

B. Scope of Homogeneous Catalysis

Homogeneous catalysis, by definition, refers to a catalytic system in which the substrates for a reaction and the catalyst components are brought together in one phase, most often the liquid phase. More recently a narrower definition has become fashionable according to which homogeneous catalysis involves organometallic complexes as the catalysts (strictly speaking an organometallic compound should contain a bond between a carbon atom and the metal, but this is not true for all catalysts to be discussed, as we will include coordination complexes in our definition). As a shorthand notation it will also be used for the present contribution, but it should be borne in mind that there are many interesting and important reactions employing homogeneous catalysts that are not organometallic or coordination complexes. Examples of such systems or catalysts are general acid and base catalysis (ester hydrolysis), organic catalysts (thiazolium ions in Cannizzarro reactions), enzymatic processes, nitrogen oxides, chloride atoms, etc.

Organometallic catalysts consist of a central metal surrounded by organic (and inorganic) ligands. Both the metal and the large variety of ligands determine the properties

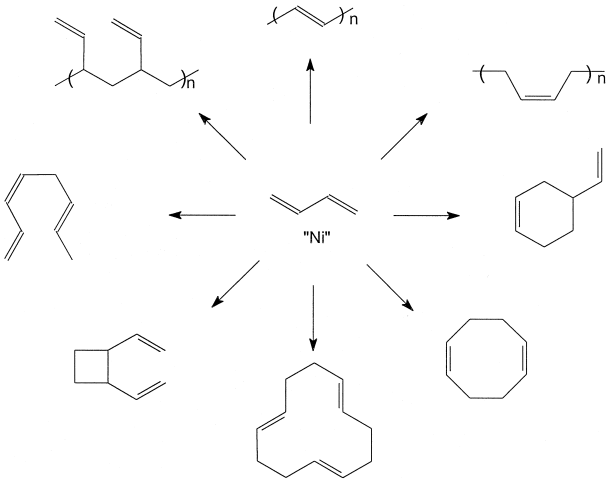


FIGURE 2 Selectivity in nickel catalyzed butadiene reactions.

of the catalyst. The success of organometallic catalysts lies in the relative ease of catalyst modification by changing the ligand environment. Crucial properties to be influenced are the acceleration of the reaction and the selectivity to certain products. One metal can give a variety of products from one single substrate simply by changing the ligands around the metal center. Figure 2 shows the products that can be obtained from butadiene with nickel catalysts.

Several types of selectivity can be distinguished in a chemical reaction: chemoselectivity; regioselectivity; diastereoselectivity; and enantioselectivity. See Fig. 3 for an explanation. Note the difference between diastereoselectivity and enantioselectivity; the former starts with a chiral substrate and an achiral catalyst to produce diastereomers, while the latter starts with an achiral (often called prochiral) substrate and the optical activity is induced by a chiral catalyst.

II. ELEMENTARY STEPS FOR HOMOGENEOUS CATALYSIS

A. Creation of a "Vacant Site" and Coordination of the Substrate

To bring the reactants together a metal center must have a vacant site. Metal catalysis begins, we could say, with the creation of a vacant site. In the condensed phase solvent molecules will always be coordinated to the metal ion and "vacant site" is an inaccurate description. Substrates are present in excess and so may be the ligands. Therefore, a competition in complex formation exists between the desired substrate and other potential ligands present in the solution. Often a negative order in one of the concentrations of the "ligands" present can be found in the expression for

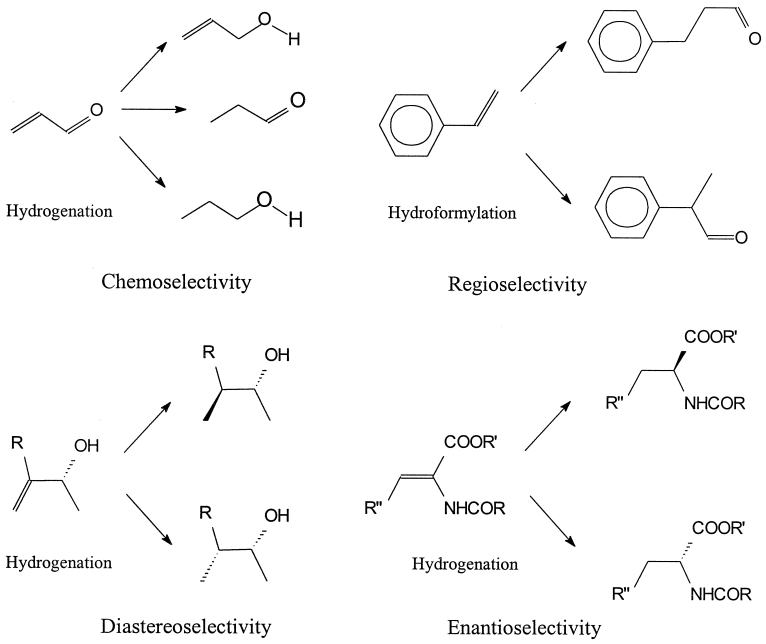


FIGURE 3 Selectivities involved in chemical conversions.

the rate of product formation. When the substrate coordinates strongly to the metal center this may give rise to a *zeroth* order in the concentration of the substrate, i.e., saturation kinetics (c.f. Michaelis-Menten kinetics as known from enzyme kinetics). Also, strong coordination of the product of the reaction may slow down or inhibit the catalytic process. These phenomena are similar to desorption and adsorption in heterogeneous catalysis. Two mechanisms are distinguished for ligand/substrate displacement, an associative and a dissociative one; see Fig. 4.

B. Insertion and Migration Reactions

Insertion and migration refer to the process in which an unsaturated molecule inserts to a metal-anion bond. Figure 5 shows how a new carbon-carbon is made when a carbon monoxide molecule inserts into a platinum methyl bond, or more accurately, when the anionic methyl group "migrates" to the unsaturated CO molecule. The empty site left by the methyl group will be occupied by another ligand, a solvent molecule in this figure.

A second important migration reaction involves alkenes instead of carbon monoxide. Figure 6 gives a schematic representation of a hydride that migrates to a coordinated ethene molecule *cis* to the hydride. The figure shows the hydride migration resulting in an empty space in the coordination sphere of the metal. This coordinative unsaturation can be lifted in two ways: first an agostic interaction with the β-hydrogens may occur and secondly an incoming ligand may occupy the vacant site.

The migration reaction of hydrides to alkenes can be described as a $2+2$ addition reaction. The reaction takes place in a *syn* fashion with respect to the alkene; the two atoms M and H add to the same face of the alkene (Fig. 7).

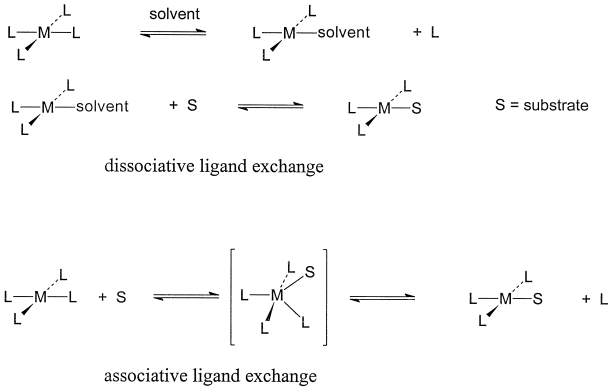


FIGURE 4 Dissociative and associative processes for ligand exchange.

C. β-Elimination and De-Insertion

The reverse reaction of the migration of η^1-bonded anionic groups to coordinated alkenes is named β-elimination

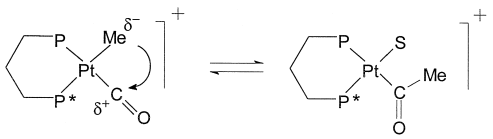


FIGURE 5 Migration reaction.

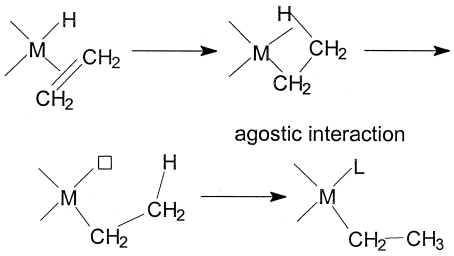


FIGURE 6 Hydride migration to ethene.

(Fig. 8). The migration reaction diminishes the total electron count of the complex by two, and creates formally a vacant site at the metal; β-elimination does the opposite. β-Elimination requires a vacant site at the complex (neglecting solvent coordination) and during the process the electron count of the complex increases by two electrons. The reaction resembles the β-elimination occurring in many organic reactions, but the difference lies in the intramolecular nature of the present process, as the eliminated alkene may be retained in the complex. In organic chemistry the reaction may well be a two-step process, e.g., proton elimination with a base followed by the leaving of the anion. In transition metal chemistry the availability of d-orbitals facilitates a concerted cis β-elimination.

Instead of β-elimination one will also find the terms de-insertion and extrusion, especially for CO. The process is completely analogous because

1. a vacant site is needed for the reaction to occur;
2. the electron count of the metal increases with two during the de-insertion (provided that the carbon monoxide remains coordinated to the metal and solvent coordination is neglected).

D. Oxidative Addition

In an oxidative addition reaction a compound XY adds to a metal complex during which the XY bond is broken and two new bonds are formed, MX and MY. X and Y are reduced and will at least formally have a minus one charge and hence the formal valency of the metal is raised by two. The coordination number of the metal increases by two. The electron count around the metal complex increases by two, but the d-electron count of the metal decreases by two. The 16-electron square planar complex is converted

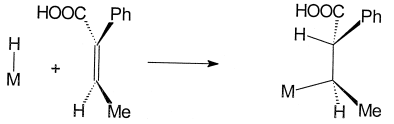


FIGURE 7 *Syn* additin of metal hydride to alkene.

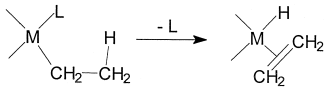


FIGURE 8 β-hydride elimination.

into an octahedral 18-electron complex. In Fig. 9 we have depicted the oxidative addition of methyl iodide to Vaska's complex (L = phosphine).

The oxidative addition of acids to zerovalent metals such as nickel is also an instructive example. It resembles the reactions with alkyl halides and may result in an "amphoteric" hydride:

$$L_4Ni + H^+ \rightarrow L_4HNi^+$$

The starting material is an 18-electron nickel zero complex which is protonated forming a divalent nickel hydride. This can react further with alkenes to give alkyl groups, but it also reacts as an acid with hard bases to regenerate the nickel zero complex.

The oxidative addition of alkyl halides can proceed in different ways, although the result is usually a trans addition independent of the mechanism. In certain cases the reaction proceeds as an S_N2 reaction as in organic chemistry. That is to say that the electron-rich metal nucleophile attacks the carbon atom of the alkyl halide, the halide being the leaving group. This process leads to inversion of the stereochemistry of the carbon atom (only when the carbon atom is asymmetric this can be observed). There are also examples in which racemization occurs. This has been explained on the basis of a radical chain mechanism. The reaction sequence for the radical chain process reads as follows:

$$L_nM + R\bullet \rightarrow L_nRM\bullet$$

$$L_nRM\bullet + RX \rightarrow L_nRMX + R\bullet$$

Oxidative additions involving C-H bond breaking is a topic of interest, usually referred to as C-H activation; the idea is that the M-H and M–hydrocarbyl bonds formed will be much more prone to functionalization than the unreactive C-H bond. Intramolecular oxidative additions of C-H bonds have been known for quite some time, see Fig. 10. This process is named orthometalation. It occurs frequently in metal complexes, and is not restricted to "ortho" protons.

Oxidative addition reactions involving carbon-to-oxygen bonds are of relevance to the catalysis with

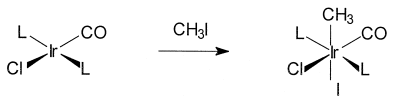


FIGURE 9 Oxidative addition of methyl iodide to Ir(I).

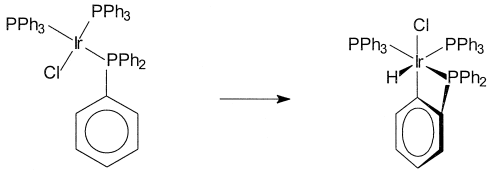


FIGURE 10 Oxidative addition of a carbon-hydrogen bond.

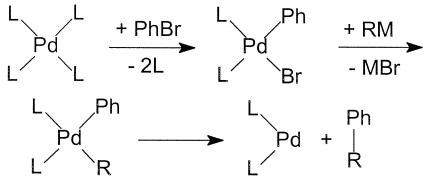


FIGURE 12 Palladium-catalyzed cross coupling.

palladium complexes. The most reactive carbon-oxygen bond is that between allylic fragments and carboxylates. The reaction starts with a palladium zero complex and the product is a π-allylic palladium(II) carboxylate (Fig. 11).

The point of interest is the "amphoteric" character of the allyl anion in this complex. On the one hand it may react as an anion, but on the other hand it is susceptible to nucleophilic attack by, for example, carbon centered anions. This has found widespread use in organic synthesis. The reaction with the anion releases a zerovalent palladium complex and in this manner palladium can be employed as a catalyst.

E. Reductive Elimination

Reductive elimination is simply the reverse reaction of oxidative addition; the formal valence state of the metal is reduced by two (or one, in a bimetallic reaction) and the total electron count of the complex is reduced by two. While oxidative addition can also be observed for main group elements, this reaction is more typical of the transition elements in particular the electronegative, noble metals. In a catalytic cycle the two reactions occur pairwise. At one stage the oxidative addition occurs, followed by, e.g., insertion reactions, and then the cycle is completed by a reductive elimination of the product.

Reductive elimination of molecules with carbon-carbon bonds has no counterpart in oxidative addition reactions. First, because the metal-carbon bonds energies may not always be large enough to compensate for the energy of the carbon-carbon bond, and secondly the carbon-carbon bond is much less reactive than a carbon-hydrogen bond or a dihydrogen bond due to repulsive interactions. In organic synthesis the palladium- or nickel-catalyzed cross-coupling presents a very common example involving oxidative addition/reductive elimination. A simplified scheme is shown in Fig. 12. The third step shows the reductive elimination.

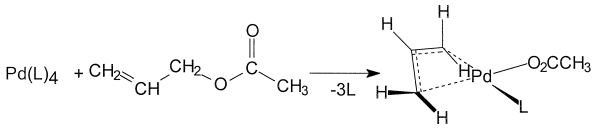


FIGURE 11 Oxidative addition of allyl acetate to Pd(0).

F. α-Elimination Reactions

α-Elimination reactions have been the subject of much study since the mid-1970s mainly due to the pioneering work of Schrock. The Early Transition Metals are most prone to α-elimination, but the number of examples of the later elements is growing. A classic example is shown in Fig. 13. The sequence of elementary steps, the nature of the reagents, and the reaction conditions are pertinent to the success of such reactions, but these details do not concern us here. Dimethyl complexes of many metals lead to formation of methane via an α-elimination process, but often the putative metal-alkylidene species is too reactive to be isolated.

Metal alkylidene complexes find application in the metathesis of alkenes, the cyclopropanation of alkenes, Wittig-type reactions, and the McMurry reaction. In suitable complexes α-elimination can occur *twice* yielding alkylidyne complexes. Fig. 14 shows a schematic example for tungsten. Alkylidyne complexes can be used as catalysts for the metathesis of alkynes.

G. Cyclometallation

Cyclometallation refers to a process of unsaturated moieties forming a metallacyclic compound. Examples of the process are presented in Fig. 15. Metal complexes revealing these reactions comprise: $M = L_2Ni$ for reaction **a**, $M = Cp_2Ti$ for reactions **b** and **c**, M = Ta for **d**, and $M = (RO)_3W$ for **e**. The latter examples involving metal-to-carbon multiple bonds have only been observed for early transition metal complexes, the same ones mentioned under α-elimination, Section F.

The reverse reaction of a cyclometallation is of importance for the construction of catalytic cycles. The retrocyclometallations of reactions **a** and **b** are not productive, unless the structures were obtained via another route. For **c**–**e** the following retro reactions can be envisaged leading to new products; see Figs. 16 and 17.

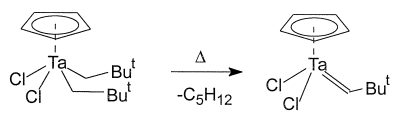


FIGURE 13 α-elimination leading to a metal alkylidne complex.

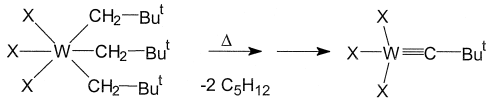


FIGURE 14 α-elimination leading to a metal alkylidyne complex.

H. Activation of a Substrate Toward Nucleophilic Attack

1. Alkenes

Coordination of an alkene to an electronegative metal activates the metal toward attack of nucleophiles. After the nucleophilic attack the alkene complex has been converted into a σ-bonded alkyl complex with the nucleophile at the β-position. With respect to the alkene (in the "organic" terminology) the alkene has undergone *anti*-addition of M and the nucleophile Nu; see Fig. 18.

As indicated under section B the overall result is the same as that of an insertion reaction, the difference being that insertion gives rise to a *syn*-addition and nucleophilic attack to an *anti*-addition. In Fig. 18 the depicted nucleophile is anionic, but Nu may also be a neutral nucleophile such as an amine. The addition reaction of this type is the keystep in the Wacker-type processes catalyzed by palladium.

2. Carbon Monoxide

Coordinated carbon monoxide is subject to activation toward nucleophilic attack. Through σ-donation and π-back donation into the antibonding CO π^* orbitals the carbon atom has obtained a positive character. This makes the carbon atom not only more susceptible for a migrating anion at the metal center, but also for a nucleophile attacking from outside the coordination sphere. In this instance it is more difficult to differentiate between the two pathways. The nucleophilic attack by alkoxides, amines, and water is of great interest to homogeneous catalysis. A dominant reaction in syn-gas systems is the conversion of carbonyls

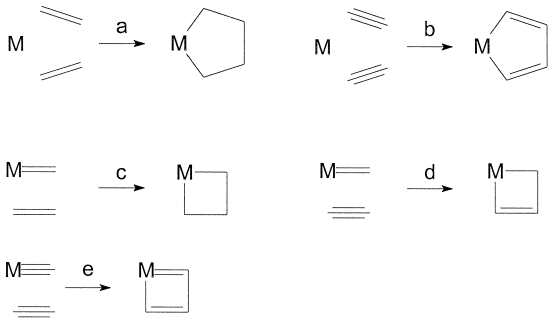


FIGURE 15 Schematic overview of cyclometallation reactions.

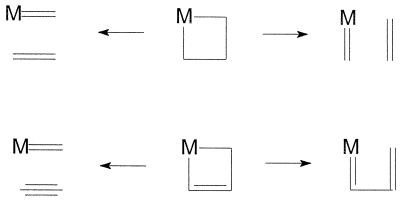


FIGURE 16 Retrocyclometallation reactions.

with water to metal hydrides and carbon dioxide ("shift reaction"); see Fig. 19.

I. σ-Bond Metathesis

A reaction that is relatively new and not mentioned in older textbooks is the so-called σ-bond metathesis. It is a concerted $2+2$ reaction immediately followed by its retrograde reaction giving metathesis. Both late and early transition metal alkyls are prone to this reaction, but its occurrence had to be particularly invoked in the case of the early transition metals; the latter often lack d-electrons needed for other mechanisms such as oxidative addition. This alternative does not exist for d^0 complexes such as Sc(III), Ti(IV), Ta(V), W(VI), etc., and in such cases σ-bond metathesis is the most plausible mechanism. In Fig. 20 the reaction has been depicted.

J. Dihydrogen Activation

In the above reactions we have not explicitly touched upon the reactions of dihydrogen and transition metal complexes. Here the reactions that involve the activation of dihydrogen will be summarized, because they are very common in homogeneous catalysis and because a comparison of the various mechanisms involved may be useful. Four reactions are usually distinguished for hydrogen:

1. oxidative addition (see Fig. 9),
2. bimetallic oxidative addition,
3. heterolytic cleavage, and
4. σ-bond metathesis (see Fig. 20).

1. Oxidative Addition

Oxidative addition of dihydrogen commonly involves transformation of a d^8 square planar metal complex into a d^6 octahedral metal complex, or similar transformations

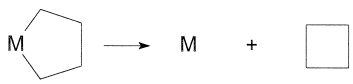


FIGURE 17 Cyclization through reductive elimination.

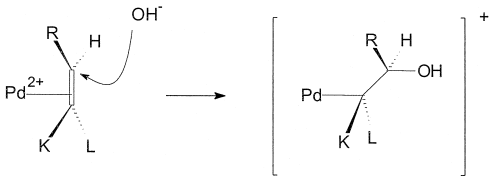


FIGURE 18 Nucleophilic *anti*-attack.

involving $d^2 \rightarrow d^0$, $d^{10} \rightarrow d^8$, etc. The oxidative addition of dihydrogen to low-valent metal complexes is a common reaction in many catalytic cycles. In spite of the high strength of the dihydrogen bond the reaction proceeds smoothly to afford cis dihydrido complexes. The bond energy of a metal hydrogen bond is in the order of 240 +/− 40 kJmol^{-1} which is sufficient to compensate for the loss of the H-H bond (436 kJmol^{-1}). The hydride is formally charged with a minus one charge and this electron count gives dihydrogen the role of an oxidizing agent. The classic example of oxidative addition to a d^8 metal complex is the reaction discovered by Vaska and Diluzio:

$$\text{trans-IrCl(CO)(PPh}_3)_2 + H_2 \rightarrow \text{IrH}_2\text{Cl(CO)(PPh}_3)_2$$

In rhodium complexes the reaction has found widespread application in hydrogenation. In model compounds the reaction reads:

$$\text{Rh(diphosphine)}_2^+ + H_2 \rightarrow \text{RhH}_2\text{(diphosphine)}_2^+$$

The second reaction, bimetallic oxidative addition—also referred to as homolytic splitting—involves a reaction of a dimeric complex with H_2 in which two metal centers participate. For instance, a dimer of a d^7 metal complex reacts with dihydrogen to give two d^6 species. In this process dihydrogen also turns formally into two hydride anions. A well-known example is the conversion of dicobaltoctacarbonyl into hydridocobalttetracarbonyl:

$$\text{Co}_2(\text{CO})_8 + H_2 \rightarrow 2\ \text{HCo(CO)}_4$$

Another example implies two molecules of a Co(II), a d^7 complex, which does suggest that radical reactions may be occurring, such as in Co(CN)_5^{-3}. Indeed in the subsequent reactions the proton is transferred as a hydrogen radical to the activated alkenes (e.g., acrylates, styrene) and the reference to the process as a homolytic dissociation is certainly applicable to the reactions with alkenes. Formally, dihydrogen is transformed into two hydride anions and the reaction is again an oxidative addition.

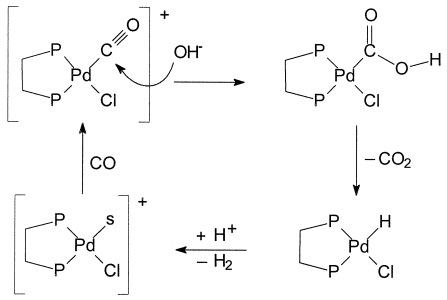


FIGURE 19 Nucleophilic attack at coordinated CO followed by a "shift" reaction.

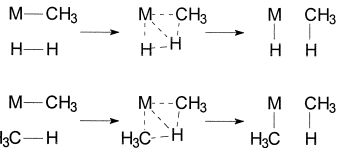


FIGURE 20 σ-Bond metathesis.

$$2\ \text{Co(CN)}_5^{-3} + H_2 \rightarrow 2\ \text{HCo(CN)}_5^{-3}$$

2. Heterolytic Splitting

Heterolytic splitting of dihydrogen refers to the splitting of dihydrogen into a proton and a metal-bonded hydride. It is a common reaction, probably for many transition metal cations, but there are only a few cases for which there is clear proof for its occurrence. In the ideal case the heterolytic splitting is catalyzed by the metal ion and a base which assists in the abstraction of the proton. In this reaction there is no formal change in the oxidation state of the metal. The mechanism has been put forward for Ru(II) complexes which can react with dihydrogen according to:

$$\text{RuCl}_2\text{(PPh}_3)_3 + H_2 \rightarrow \text{RuHCl(PPh}_3)_3 + \text{HCl}$$

Ruthenium has a sufficient number of d-electrons to allow for oxidative addition of dihydrogen, which could then be quickly followed by reductive elimination of HCl. Observations on dihydrogen complexes of ruthenium have thrown light on the mechanism heterolytic splitting. $\text{CpRu(L)}((\eta^2\text{-H}_2)^+$ reacts rapidly with NEt_3 as can be deduced from the dynamic ^{1}H-NMR spectra which indicate a rapid exchange of the dihydrogen complex with its conjugate base, CpRu(L)(H) (Fig. 21).

K. Activation by Lewis Acids

In Section H we have discussed the activation of carbon containing fragments toward nucleophilic attack by coordination of the fragment to a transition metal. Here we will

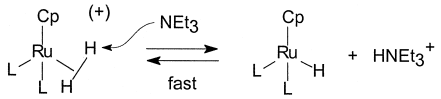


FIGURE 21 η^2-H_2 complex and heterolytic splitting.

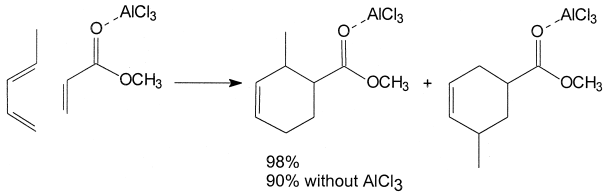


FIGURE 22 Lewis acid catalyzed Diels-Alder reaction.

describe a few examples of activation of reagents when complexed to Lewis acids. In organic textbooks one will find a variety of reactions catalyzed by Lewis acids.

1. Diels-Alder Additions

The Diels-Alder reaction is the reaction of a diene with a mono-ene to form a cyclohexene derivative, an important reaction for the construction of organic intermediates. One of its attractions is the atom efficiency of 100% with no by-products being formed. The mono-ene, or dienophile which may also be an alkyne, has a LUMO of low energy while the diene is usually electron rich with a high-lying HOMO. The interaction of these two orbitals starts the reaction between the two molecules (Fig. 22).

The reaction can be accelerated by complexation of the dienophile to a Lewis acid, which further lowers the level of the interacting LUMO: two regioisomers are formed in the example shown, we shall not consider the stereochemistry. Typically in these systems governed by frontier orbitals, the reactions not only become much faster with a Lewis acid catalyst, but also more regioselective, that is to say the pathways to different isomers may experience different accelerations.

2. Epoxidation

Alkenes can be transformed into epoxides by hydroperoxides and a catalyst, which often is a high-valent titanium or molybdenum complex acting as a Lewis acid. The mechanism is not clear in great detail; in Fig. 23 a suggested mechanism is given. The key factor is the action of the metal on the peroxo group making one oxygen atom electrophilic.

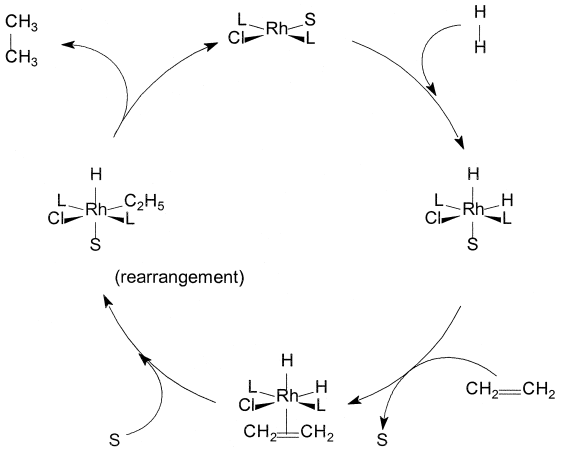


FIGURE 24 Wilkinson's hydrogenation cycle.

3. Ester Condensation

An application of industrial importance of Lewis acidic metal salts is the condensation of carboxylic diacids and diols to give polyesters. This is an acid catalyzed reaction that in the laboratory is usually catalyzed by protic acids. For this industrial application salts of manganese, nickel, or cobalt and the like are used.

III. DESCRIPTION AND MECHANISMS OF HOMOGENEOUS CATALYTIC PROCESSES

A. Hydrogenation

1. Wilkinson's Catalyst

Undoubtedly the most popular homogeneous catalyst for hydrogenation is Wilkinson's catalyst, $RhCl(PPh_3)_3$, discovered in the 1960s. The reaction mechanism, its dependence on many parameters, and its scope have been studied in considerable detail.

The catalytic cycle is shown in Fig. 24. Note that the reactions have been drawn as irreversible reactions while most of them are actually equilibria. In this scheme L stands for triarylphosphines and S for solvent (ethanol, toluene). The alkene is simply ethene. For simplicity we

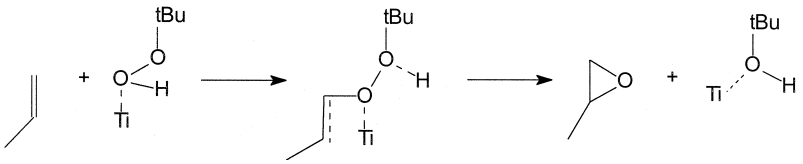


FIGURE 23 Epoxidation catalyzed by Lewis acids.

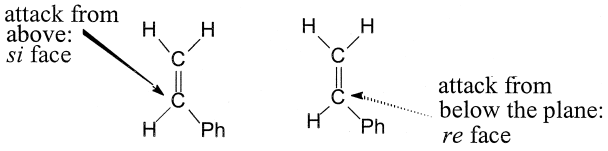


FIGURE 25 *Re*- and *si*-faces on an alkene substituted at one atom.

have chosen Wilkinson's complex as the starting point in our cycle. As in many other cases to follow the number of valence electrons switches during the cycle between two numbers differing by two electrons; in this instance the valence electron counts switch between 16 and 18. A 14-electron count for the unsaturated species occurring at the beginning of the cycle has also been discussed. The first step in this sequence is the dissociation of one ligand L that is replaced by a solvent molecule.

After ligand dissociation an oxidative addition reaction of dihydrogen takes place. As usual this occurs in *cis* fashion and can be promoted by the substitution of more electron-rich phosphines on the rhodium complex. The next step is the migration of hydride forming the ethyl group. Reductive elimination of ethane completes the cycle. Obviously, employing electron-withdrawing ligands can increase the rate of this step. Within a narrow window of aryl phosphines small changes in rates have been observed which could indeed be explained along the lines sketched above. Strong donor ligands, however, stabilize the trivalent rhodium(III) chloro dihydride to such an extent that the complexes are no longer active.

2. Asymmetric Hydrogenation

a. Introduction "prochirality." Planar molecules possessing a double bond such as alkenes, imines, and ketones, which do not contain a chiral carbon in one of the side chains, are not chiral. When these molecules coordinate to a metal a chiral complex is formed, unless the alkene possesses C_{2V} symmetry. A simple silver cation Ag^+ suffices. In other words, even a simple alkene such as propene will form a chiral complex when it coordinates to a transition metal, so will trans-2-butene, but cis-2-butene won't. If a bare metal atom coordinates to cis-2-butene the complex has a mirror plane, and hence the complex is not chiral. In summary, with few exceptions, complexation of a metal to the one face of an alkene gives rise to a certain enantiomer and complexation to the other face gives rise to the other enantiomer.

For complexation of planar molecules to metals rules have been introduced to allow us to denote the faces of the planar molecule; they are called the *re*-face and the *si*-face. Usually this simple annotation takes into account that only one carbon atom is used. It may be more complicated when the two carbon atoms of the alkene give rise to two stereocenters. In Fig. 25 we have drawn how we can distinguish the two faces of a simple alkene, or rather the side of attack of a specific atom of the alkene.

When a metal complex is chiral, either because it contains a chiral ligand or a chiral metal center, and then forms a complex to a "prochiral alkene," the resulting complex is a diastereoisomer. Thus, a mixture of diastereomers can form when the chiral complex coordinates to each face of the alkene. As usual, these diastereomers have different properties and can be separated. Or, more interestingly for catalysis, the two diastereomers are formed in different amounts.

b. Enantioselective hydrogenation. The asymmetric hydrogenation of cinnamic acid derivatives has been developed by Knowles at Monsanto. The synthesis of L-dopa (Fig. 26), a drug for the treatment of Parkinson's disease, has been developed and applied on an industrial scale. The reaction is carried out with a cationic rhodium complex and an asymmetric diphosphine as the ligand that induces the enantioselectivity. Surprisingly, the reaction is not very sensitive to the type of diphosphine used, although it must be added that most ligands tested are bis(diphenylphosphino) derivatives. On the other hand, the reaction is very sensitive to the type of substrate and the polar substituents are prerequisites for a successful asymmetric hydrogenation. Fig. 26 shows the overall reaction.

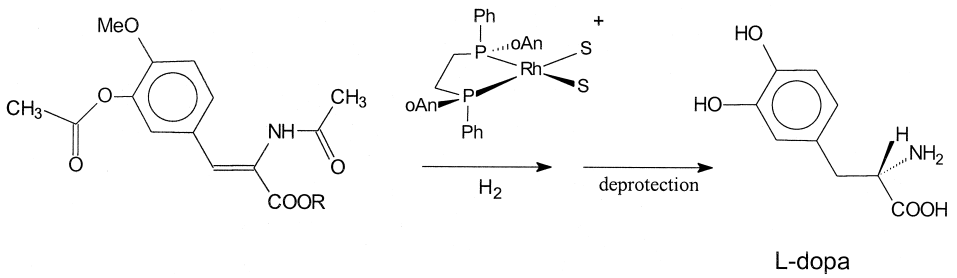


FIGURE 26 Synthesis of L-dopa (oAn = ortho-anisyl = 2-methoxyphenyl).

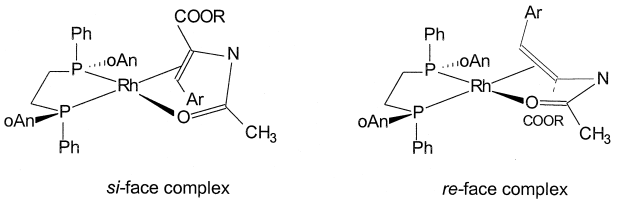


FIGURE 27 Diastereomeric alkene complexes.

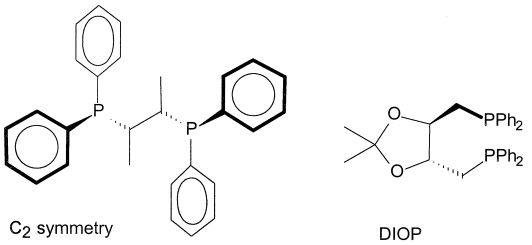


FIGURE 29 C_2-Chiral diphosphines.

The hydrogenation reaction is carried out with a substituted cinnamic acid. The acetamido group is of particular importance because it functions as a secondary complexation function in addition to the alkene functionality. In the first step the alkene coordinates to the cationic rhodium species for which there are two possibilities, binding to the *re*-face and the *si*-face (Fig. 27). This new chiral center, combined with the ligand chirality, leads to the formation of two diastereomers which have different free energies and reactivities. Thus, a preference for just one enantiospecific pathway may result. For a few systems *in situ* characterizations have been carried out and both complexes have been identified. The final product is not necessarily derived from the most abundant isomer; all that matters is through which pathway does the fastest reaction take place that produces the most product.

The formation of the *re* and *si* adducts is reversible and thus in this step no chirality is determined yet. As soon as an irreversible step occurs—after chirality has been introduced—the chirality of the final product has been determined. In the present reaction this step may be either the oxidative of addition of H_2 or the migratory insertion (Fig. 28).

A large series of asymmetric ligands have been developed most of which have the asymmetric "center" in the bridge rather than at the phosphorus atoms as is the case in DIPAMP, the ligand shown in Fig. 28 for the L-dopa synthesis. Their synthesis is similar to the one developed by Kagan for DIOP (see Fig. 29) starting from asymmetric acids which can be commercially obtained (tartaric acid in the case of DIOP). Crucial to the success of this family of ligands is the C_2 "propeller"-type symmetry which divides the space around rhodium (or any metal) into four quadrants, two relatively empty and two filled ones (see Fig. 29).

Two phenyl groups at one phosphorus of the chelating ring adopt an axial and an equatorial position. This explains the similarity of this large family of ligands in the enantioselective reactions. Coordination of the dehydroalanine derivative (or enamides in general) will now take place in such a way that the auxiliary donor atom coordinates to one site, and the phenyl-substituted alkene will coordinate to the other site with the face that gives the least interaction of its substituents with the phenyl group of the ligand pointing into the same quadrant. This gives the predominant metal alkene adduct.

Three other ways of achieving chirality in phosphines should be mentioned. The first one also concerns C_2 chiral ligands of the type DuPHOS (Fig. 30) developed by Burk.

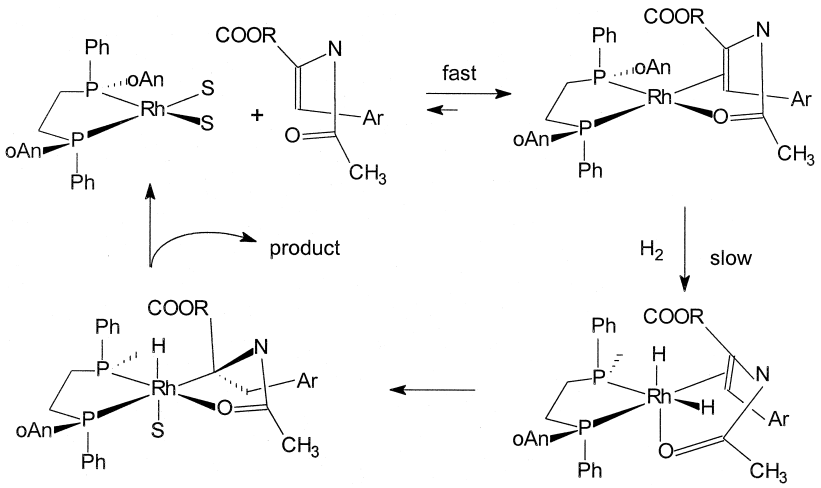


FIGURE 28 Asymmetric hydrogenation.

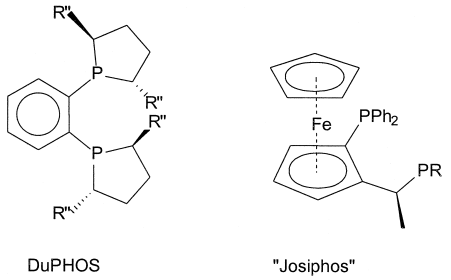


FIGURE 30 Chiral diphosphine ligands.

The chirality resides neither on the backbone nor on the phosphorus atom, but on the cyclic substituents. Modification of the R-groups leads to highly efficient catalysts. C_1-chiral ligands can also be very selective as has been shown by Togni who developed a range of ligands, here exemplified by Josiphos. The ligand contains "two" chiral centers, one at the carbon atom and the other involves facial chirality of the cyclopentadienyl plane. Derivatives of this type are used for the commercial production of chiral pharmaceuticals and agrochemicals. Another way to achieve chirality is by hindered rotation around molecular axes ("axial" chirality) as will be shown in the next section where BINAP will be introduced. Application DIOP?

B. Isomerization

1. Insertion and β-Elimination

A catalytic cycle which involves only one type of elementary reaction must be a very facile process. Isomerization is such a process since only migratory insertion and its counterpart β-hydride elimination are required. Hence the metal complex can be optimized to do exactly this reaction as fast as possible. The actual situation is slightly more complex due to the necessity of vacant sites, which have to be created for alkene complexation and for β-elimination. As expected, many unsaturated transition metal hydride complexes catalyze isomerization. Examples include monohydrides of Rh(I), Pd(II), Ni(II), Pt(II), and Zr(IV). The general scheme for alkene isomerization is very simple; for instance it may read as shown in Fig. 31.

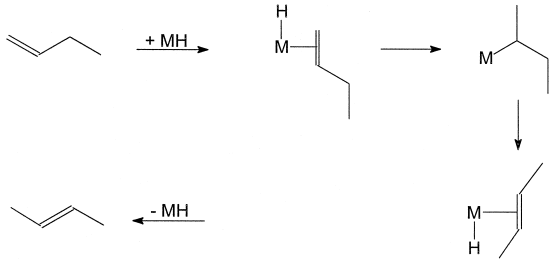


FIGURE 31 Simplified isomerization scheme involving β-hydride elimination.

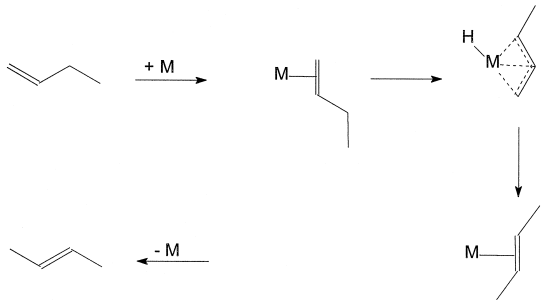


FIGURE 32 Simplified allylic isomerization scheme.

2. Allylic Mechanism

A second mechanism that has been brought forward involves the formation of allylic intermediates. The presence of a hydride on the metal complex is not required in this mechanism which can best be described as an oxidative addition of an "activated" C-H bond (i.e. an allylic hydrogen) to the metal. The allyl group can recollect its hydrogen at the other end of the allyl group and the result is also a 1,3 shift of hydrogen (Fig. 32).

3. Asymmetric Isomerization

An important application of an isomerization is found in the Takasago process for the commercial production of (−)menthol from myrcene. The catalyst used is a rhodium complex of BINAP. The BINAP complex is an asymmetric ligand based on the atropisomerism of substituted dinaphthyl (Fig. 33). It was first introduced by Noyori. Atropisomers of diphenyl and the like are formed when ortho substituents do not allow rotation around the central carbon-carbon bond. As a result two enantiomers are formed.

For asymmetric hydrogenation, transfer hydrogenation, and isomerization of double bonds using both ruthenium and rhodium complexes BINAP has been extensively used. The synthesis of menthol is given in the reaction scheme, Fig. 34. The key reaction is the enantioselective isomerization of the allylamine to the asymmetric enamine. It is proposed that this reaction proceeds via an allylic intermediate.

This is the only step that needs to be steered to the correct enantiomer, since the other two are produced in the

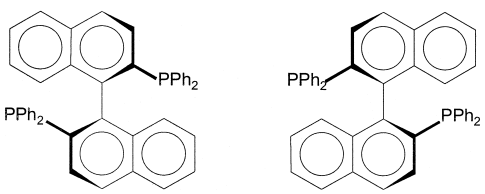


FIGURE 33 The two enantiomers of BINAP.

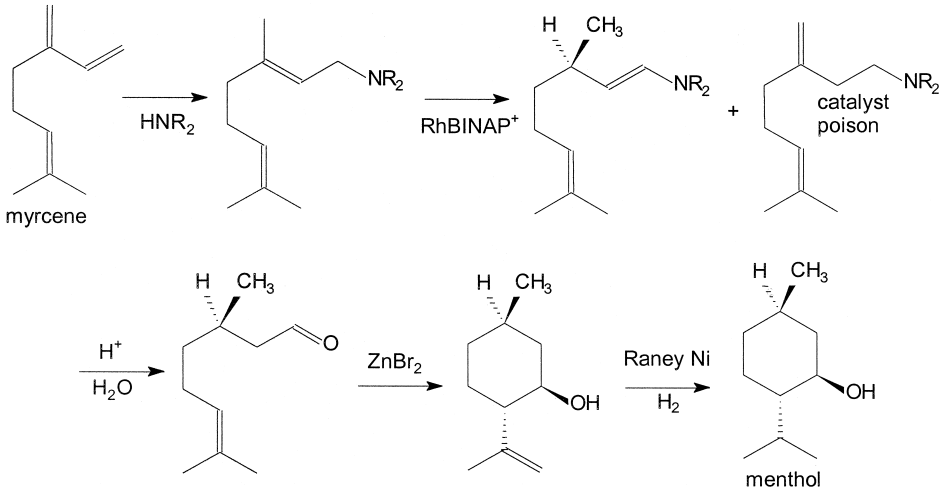


FIGURE 34 The Takasago process for (−)menthol.

desired stereochemistry with the route depicted. Of the eight possible isomers only this one (1R,3R,4S) is important. After the enantioselective isomerization the enamine is hydrolyzed. A Lewis acid catalyzed ring closure gives the menthol skeleton. In a subsequent step the isopropenyl group is hydrogenated over a heterogeneous Raney nickel catalyst. Asymmetric catalysis involving metal-catalyzed hydrogenations and isomerization is becoming increasingly important in the production of pharmaceuticals, agrochemicals, and flavors and fragrances.

C. Carbonylation of Methanol

1. Monsanto Acetic Acid Process

The carbonylation of methanol has been developed by Monsanto in the late 1960s. It is a large-scale operation employing a rhodium/iodide catalyst converting methanol and carbon monoxide into acetic acid. Since the 1990s an iridium/iodide-based catalyst has been used by BP.

$$CH_3OH + CO \rightarrow CH_3COOH$$

$$\Delta G, \text{ standard conditions}, -17.8 \text{ kcal/mol}$$

The same carbonylation reaction can be carried out with a cobalt catalyst. The rhodium and iridium catalysts have several distinct advantages over the cobalt catalyst; they are much faster and far more selective. The higher rate is in process terms translated into much lower pressures (the cobalt catalyst is operated at pressures of 700 bar).

For years the Monsanto process has been the most attractive route for the preparation of acetic acid. The two components are rhodium and iodide, which can be added in many forms. A large excess of iodide may be present, and both on a weight basis as well as on a molecular basis, it would be fair to say that the catalyst is iodide and the promoter is a trace of rhodium. Methanol and the iodide component form under the reaction conditions methyl iodide. Rhodium is present as the anionic species $RhI_2(CO)_2^-$. The first step of the catalytic cycle (see Fig. 35) is the oxidative addition of methyl iodide to this rhodium complex. Migration of the methyl group gives an acetyl rhodium complex. CO complexation and reductive elimination of acetyl iodide complete the cycle. Acetyl iodide hydrolyzes to give acetic acid and hydrogen iodide. The latter regenerates methyl iodide in a reaction with methanol.

It is important that in the two "organic" equilibria involving iodide the one with methanol involves complete conversion to methyl iodide, whereas acetyl iodide is completely converted into acetic acid and hydrogen iodide:

$$CH_3OH + HI \longrightarrow H_2O + CH_3I$$

$$CH_3COI + H_2O \longrightarrow CH_3CO_2H + HI$$

The rate-determining step in this process is the oxidative addition of methyl iodide. The iodide enables the formation of a methyl rhodium complex; methanol is not sufficiently electrophilic to carry out this reaction. Within the operation window of the process the reaction rate is independent of the carbon monoxide pressure. Furthermore, the methyl iodide formation from methanol is almost complete, which makes the reaction rate also practically independent of the methanol concentration. As a result of the kinetics and the equilibria mentioned above, all iodide in the system occurs as methyl iodide. Hence, up to high conversions the rate does not depend on the

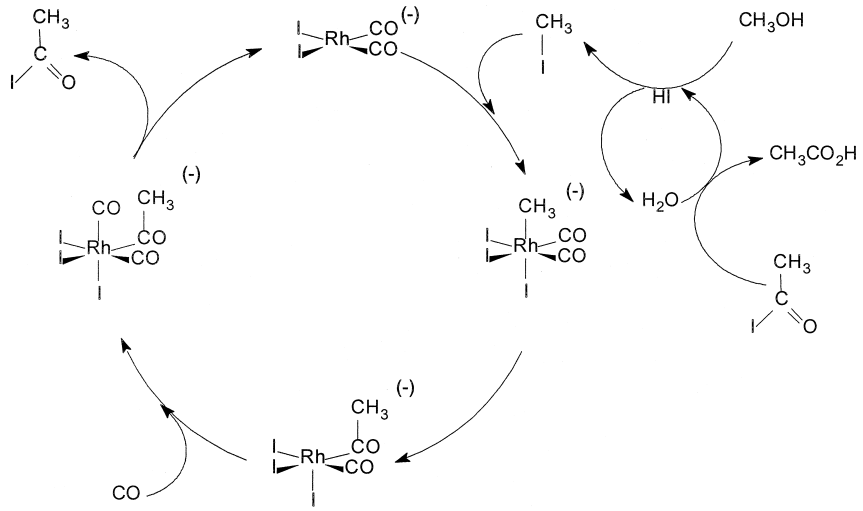


FIGURE 35 Monsanto carbonylation of methanol.

concentrations of the two reactants. The process scheme is presented in Fig. 36.

2. Acetic Anhydride

In many applications acetic acid is used as the anhydride and the synthesis of the latter is therefore equally important. In the 1970s Halcon (now Eastman) and Hoechst (now Celanese) have developed a process for the conversion of methyl acetate and carbon monoxide to acetic anhydride. The process has been on stream since 1983 and several 100,000 are being produced annually together with some 10–20% acetic acid. The reaction is carried out under similar conditions as the Monsanto process, and also uses methyl iodide as the "activator" for the methyl group. The reaction scheme follows that of the Monsanto process except for the "organic" cycle, in which acetic acid replaces water, and methyl acetate replaces methanol (Fig. 35):

$$CH_3COOCH_3 + HI \rightleftharpoons CH_3COOH + CH_3I$$

$$CH_3COI + CH_3COOH \rightleftharpoons CH_3COOOCH_3 + HI$$

The first reaction generates methyl iodide for the oxidative addition, and the second reaction converts the reductive elimination product acetyl iodide into the product and it regenerates hydrogen iodide. There are, however, a few distinct differences between the two processes. The thermodynamics of the acetic anhydride formation are less favorable and the process is operated much closer to equilibrium. Under standard conditions the ΔG values are approximately:

$$MeOAc + CO \rightarrow Ac_2O \qquad -2.5 \text{ kcal/mol}$$

Two more differences are

1. in the Eastman process 5% of H_2 is added to the carbon monoxide to accomplish reduction of trivalent rhodium, and
2. the addition of cations such as Li^+ or Na^+ is necessary.

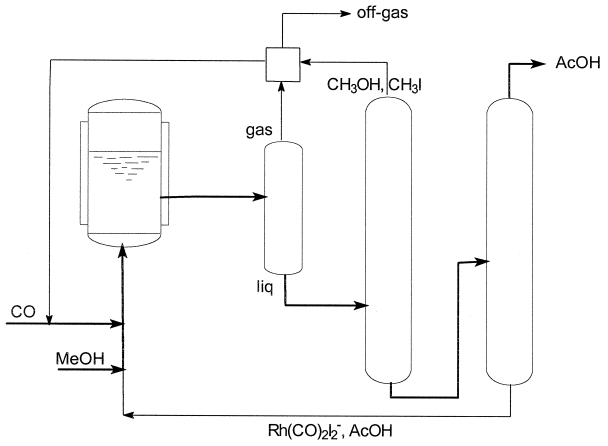


FIGURE 36 Simplified flow-scheme of the Monsanto process for acetic acid.

D. Hydroformylation

1. Introduction

Functionalization of hydrocarbons from petroleum sources is mainly concerned with the introduction of oxygen into the molecule. Roughly speaking, two ways are open: oxidation and carbonylation. Oxidation is the preferred route *inter alia* for aromatic acids, acrolein, maleic anhydride, ethene oxide, propene oxide, and acetaldehyde. Hydroformylation (older literature and

technical literature say "oxo" reaction) is employed for the large-scale preparation of butanal and butanol, 2-ethyl hexanol, and detergent alcohols. Butanal and butanol are used in many applications as a solvent, in esters, in polymers, etc. The main use of 2-ethylhexanol is in phthalate esters, which are softeners (plasticizers) in PVC. The catalysts applied are based, again, on cobalt and rhodium.

2. Cobalt-Based Oxo-Process

Roelen accidentally discovered the hydroformylation of alkenes in the late 1930s while he was studying the conversion of synthesis gas to liquid fuels (Fischer-Tropsch reaction) using a heterogeneous cobalt catalyst. It took more than a decade before the reaction was taken further, but now it was the conversion of petrochemical hydrocarbons into oxygenates that was the driving force. It was discovered that the reaction was not catalyzed by the supported cobalt but, in fact, by $HCo(CO)_4$ formed in the liquid state.

A key issue in the hydroformylation reaction is the ratio of "normal" (linear) and "iso" (branched) product being produced. Figure 37 explains this colloquial expression. The linear ("normal") product is the desired product; the value of butanal is higher because this is the product which can be converted to 2-ethyl hexanol via a base-catalyzed aldol condensation and a hydrogenation. The detergent alcohols should be preferably linear because their biodegradability was reported to be better than that of the branched product. The linearity obtained in the cobalt-catalyzed process is 60–80%. The reaction mechanism for cobalt is similar to that of rhodium, which will be discussed in the next section.

Thermodynamics of hydroformylation and hydrogenation at standard conditions are as follows:

$$H_2 + CH_3CH{=}CH_2 + CO \rightarrow CH_3CH_2CH_2C(O)H$$

ΔG	15	−33	−28(l)	= −10 kcal/mol
ΔH	5	−26	−57	= −36 kcal/mol

$$H_2 + CH_3CH{=}CH_2 \rightarrow CH_3CH_2CH_3$$

ΔG	15	−6	= −21 kcal/mol
ΔH	5	−25	= −30 kcal/mol

Thus the reaction is highly exothermic and favored by thermodynamics at temperatures roughly below 200°C. Hydrogenation of the alkene to alkane is thermodynamically even more attractive. Often this reaction is observed as a side reaction.

In industrial practice the older cobalt catalyst is still used today for the conversion of higher alkenes to detergent aldehydes or alcohols ($>C_{12}$). The cobalt process requires high pressures (70–100 bar) and temperatures (140–170°C). Aldol condensation and hydrogenation of the alkene to alkane (∼10%) are undesirable side reactions for the detergent alcohols. Interestingly, the cheaper internal alkenes can be used for this process and yet the outcome is mainly a terminally hydroformylated, linear aldehyde. Often the separation of catalyst, product, by-product, and starting material is tedious. For propene the most economic processes are rhodium-based catalysts commercialized in the 1970s.

3. Rhodium-Based Hydroformylation

Fundamental work by Nobel laureate Wilkinson demonstrated that rhodium triphenylphosphine catalysts allowed the operation of the hydroformylation reaction at much lower pressure (1 bar was reported by Wilkinson) and temperature than the cobalt process. The selectivity was also reported to be considerably higher, virtually no hydrogenation was observed and the linearity was in some cases as high as 95%. The rhodium catalysts were reported to be three orders of magnitude faster in rate. The resulting milder reaction conditions would give much less condensation products. In 1971 Union Carbide Corporation, Johnson and Matthey, and Davy Powergas (now Kvearner) joined forces to develop a process based on this new finding. As yet it is only applied for propene. Hydroformylation of propene is of prime importance and worldwide probably more than 7 million tons of butanal are produced this way annually.

A convenient catalyst precursor is $RhH(CO)(PPh_3)_3$. Under ambient conditions it will slowly convert 1-alkenes into the expected aldehydes. Internal alkenes exhibit hardly any reaction. At higher temperatures (100–120°C) pressures of 10–30 bar are required. Unless a large excess of ligand is present the catalyst will also show some isomerization activity, but the internal alkenes thus formed

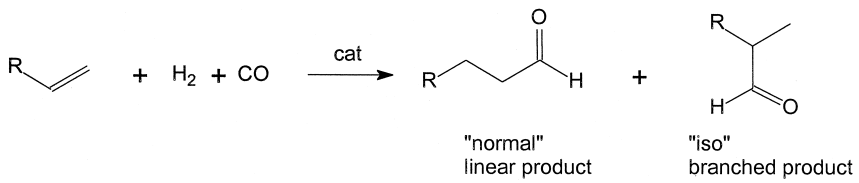


FIGURE 37 The hydroformylation reaction.

will not be hydroformylated. The 2-alkene concentration will increase and the 1-alkene concentration will decrease; this will slow down the rate of the hydroformylation reaction. This makes the rhodium catalyst less suited for the conversion of alkenes other than propene for which isomerization is irrelevant. To date, hydroformylation of higher alkenes is industrially still carried out with cobalt catalysts.

Propene hydroformylation can be done yielding a linearity ranging from 60 to 95% dependent on the phosphine concentration. At very high phosphine concentration the rate is low, but the linearity achieves its maximum value. The commercial process operates presumably around 30 bar of syn-gas, at 120°C, at high phosphine concentrations, and linearities around 92%. The estimated turnover frequency of moles of product per mole of rhodium complex per hour is in the order of 300. Low ligand concentrations, with concomitant low linearities, will give turnover frequencies in the order of 10,000 at 10 bar and 90°C. In the presence of carbon monoxide this rhodium catalyst has no activity for hydrogenation and the selectivity based on starting material is virtually 100%. The *n*-butanal produced contains no alcohol and can be converted both to butanol, to 2-ethyl-hexanol-1, and to other products as desired.

The most likely mechanism for the reaction is given in Fig. 38. Note the trigonal bipyramidal structure (tbp) of the rhodium catalysts. The σ-bonded hydrido group or the alkyl group are bound in an *apical* position of the tbp structure. This leaves three equatorial and one apical position for the remaining four ligands, carbon monoxide, phosphorus ligands, and the alkene substrate. As usual, the ligand replacement equilibria have been simplified. After an insertion step a 16-electron species is formed which restores its 18-electron count with a new CO molecule. There is consensus in the literature that the catalytic species with two phosphine ligands plays a major role, certainly when L = PPh_3. The high preference for linear aldehyde formation is ascribed to steric factors—congestion at the rhodium center favors the formation of linear alkyl and acyl species. Especially when two phosphine ligands coordinate bis-equatorially, there is a strong preference for the formation of *n*-alkyl groups. Later we will see how one can make use of this concept to obtain highly selective catalysts.

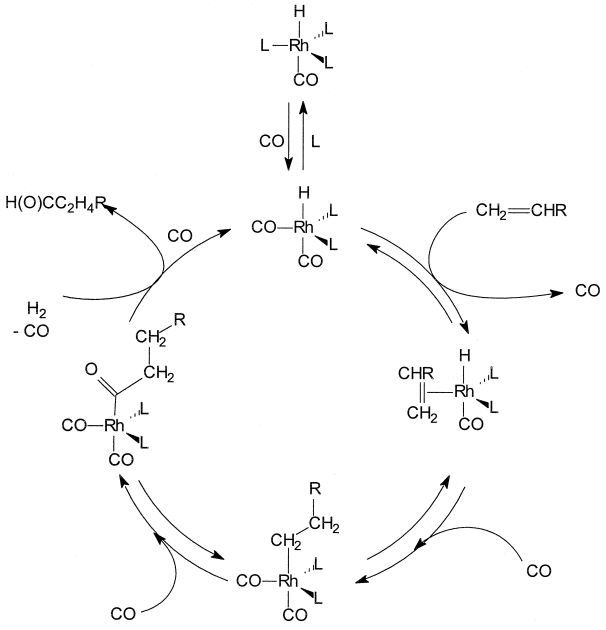


FIGURE 38 Rhodium-catalyzed hydroformylation.

For rhodium-phosphine catalysts, the kinetics shows that one of the first steps, complexation of alkene or migratory insertion of the alkene, is the rate-determining step. A first order in alkene is observed, while there is a negative dependence in the CO and/or phosphine concentration, which also stems from the replacement reaction. The rate of reaction is independent of the hydrogen concentration.

Another limiting case is the hydroformylation using electron-poor rhodium catalysts (e.g., $HRh(CO)_4$ or $HRh(CO)_3L$, L being an electron-poor phosphite). In this instance the oxidative addition of dihydrogen is the slowest step of the cycle and now the reaction shows a negative order in CO pressure, a first-order dependency in H_2, and the reaction rate is independent of alkene concentration, i.e., saturation kinetics with respect to alkene. Often the expression is more complicated and does not reveal the presence of a single slow step in the process.

4. Ligand Effects

Phosphine-based ligands have found widespread application not only in organometallic chemistry but also in industrial applications of homogeneous catalysis. Their steric and electronic properties have been widely studied to establish structure-activity relationships. The σ-basicity and π-acidity of phosphorus ligands can be compared by looking at the stretching frequencies of the coordinated carbon monoxide ligands in complexes such as $NiL(CO)_3$ where L is the phosphorus ligand. Strong σ–donor ligands give a high electron density on the metal and hence a substantial back-donation to the CO ligands and lowered IR frequencies. Strong π–acceptor ligands will compete with CO for the electron back-donation and the CO stretch frequencies will remain high. The IR frequencies represent a reliable yardstick of the electronic properties of a series of phosphine ligands toward a particular metal.

Thus, the electronic parameter for ligands with the same donor atom can be fairly well measured and applied in a

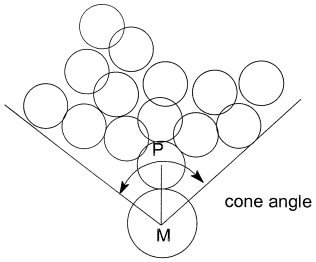


FIGURE 39 Tolman's cone angle.

qualitative sense. Many attempts have been undertaken to define a reliable steric parameter complementary to the electronic parameter. Most often used is Tolman's parameter θ (theta). Tolman proposed to measure the steric bulk of a phosphine ligand from CPK models in the following way. From the metal center, located at a distance of 2.28 Å from the phosphorus atom in the appropriate direction, a cone is constructed which embraces all the atoms of the substituents on the phosphorus atom (see Fig. 39).

The cone angle is measured, and these cone angles θ (simply in degrees) are the desired steric parameters. Crystal structure determinations have shown that in practice the angles realized in the structures are smaller than the θ-values would suggest. In reality, intermingling of the R-substituents leads to smaller effective cone angles.

Some typical values are:

Ligand PR_3, R=:	θ-value:
H	87
CH_3O	107
n-Bu	132
PhO	128
Ph	145
i-Pr	160
C_6H_{11}	170
t-Bu	182

5. Ligand Effects in Hydroformylation

Ligand effects on the selectivity of the rhodium-catalyzed hydroformylation has been extensively studied. The rate may vary several orders of magnitude and also the selectivity to either the branched or the linear product may vary dramatically with minor changes in the electronic and steric properties of the ligand.

6. Bidentates with Constrained Bite Angles

A few classes of new ligands have been introduced that give very high selectivity to linear products. All use the concept that a bis-equatorial coordination of the phosphorus ligands is needed to raise the selectivity to linear product. As mentioned above for triphenylphosphine, it was thought that also in this instance this coordination mode gave rise to the desired selectivity. The two structures are shown in Fig. 40.

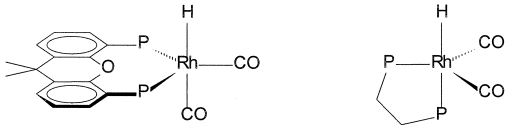


FIGURE 40 Preferred coordination modes of Xantphos (left) and dppe.

Bidentate ligands have been made and used extensively in many studies on coordination compounds and organometallic compounds. The majority of bidentate ligands, however, lead to "bite" angle of the bidentate on the metal between 75 and 95°. If a bidentate is to coordinate in a bis-equatorial fashion, it should have a bite angle of ~120°. Only very few ligands are available. Three of them are shown in Fig. 41. Molecular modeling and crystallographic studies by Casey and van Leeuwen have learned that indeed the "natural" bite angle of these ligands is 110–120°. They give linear to branched product ratios up to 100.

7. Two-Phase Hydroformylation: Water-Soluble Catalysts

In the 1980s a new process has come on stream employing a two-phase system with rhodium in a water phase and the substrate and the product in an organic phase. The catalyst used is a rhodium complex trisulfonated triphenylphosphine (tppts) (Fig. 42) which is highly water-soluble (in the order of 1 kg of the ligand "dissolves" in 1 kg of water). The ligand forms complexes with rhodium that are very similar to the ordinary triphenylphosphine complexes (i.e., $RhH(CO)(PPh_3)_3$). The rhodium complex resides in the aqueous phase. The substrates, propene/CO/H_2, are very slightly water-soluble. Upon reaction a second phase forms, the organic phase, consisting of the product butanal and dissolved propene and syn-gas. The two phases are intensely stirred to ensure a fast transport of the gases from the gas phase to the organic and water phase.

The two phases are removed from the reactor and separated in a settler tank after the heat and the gases have been removed. Often in organic synthesis the separation of two layers is not such an easy process step, but in this particular instance a very clean separation of the two layers occurs. Most importantly, the rhodium/ligand components remain completely in the aqueous phase. This type of separation is very attractive in a large-scale, continuous process. For the flow-scheme of the process, see Fig. 43.

Ruhrchemie has commercialized the process, after the initial work had been done by workers at Rhone-Poulenc,

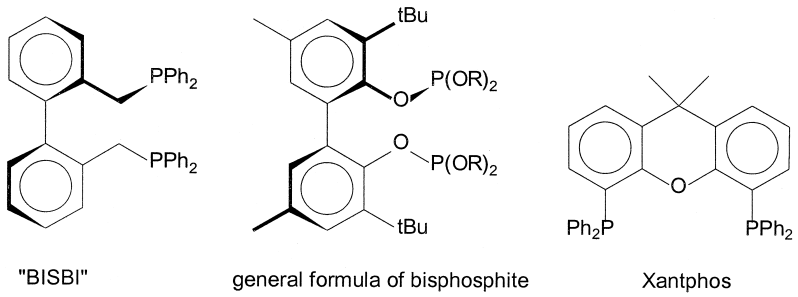


FIGURE 41 Ligands leading to high l/b ratios in rhodium-catalyzed hydroformylation.

for the production of butanal from propene. The linearity of the product amounts to 92%. The process is not applicable to the hydroformylation of higher alkenes because

- the isomerization cannot be completely suppressed with aryl phosphines, thus leading to the formation of 2-alkenes which cannot be separated from the 1-alkenes in the recycle,
- and the solubility of higher alkenes in water is very low.

E. Alkene Oligomerization

1. Introduction

1-Alkenes, or linear α-olefins as they are called in industry, are desirable starting materials for a variety of products. Polymers and detergents are the largest end uses. We mention a few applications:

C_4	Polybutylene
C_{6-8}	Comonomers in HDPE, LLDPE, synthetic esters
C_{6-10}	Alcohols (hydroformylation) as phthalates for PVC plasticizers
C_{8-10}	As trimers in synthetic lub-oils
C_{10-14}	After hydroformylation, detergents
C_{14-16}	Sulfates and sulfonates in detergents

Industrially, alkenes are obtained from several reactions, one being ethene oligomerization. Three processes are available, two based on aluminum alkyl compounds or catalysts and one on nickel catalysts.

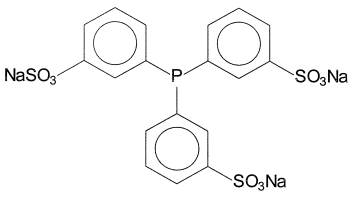


FIGURE 42 Sulfonated triphenylphosphine (tppts).

2. Shell-Higher-Olefins-Process

Many transition metal hydrides will polymerize ethene to polymeric material or, alternatively, dimerize it to butene. Fine-tuning of these catalysts to one that will give a mixture of, for example, predominantly C_{10} to C_{20} oligomers is not at all trivial. Nickel complexes have been extensively studied by Wilke and his coworkers for their activity as alkene oligomerization catalysts. In the late 1960s Keim and coworkers at Shell discovered a homogeneous nickel catalyst which selectively oligomerizes ethene to higher homologs (Fig. 44).

a. Oligomerization. The catalyst is prepared in a prereactor from nickel salts with boron hydrides as the reductor under a pressure of ethene and then ligand is added. Polar solvents such as alcohols are used for the dissolution of the catalyst. The catalyst solution and ethene are led to the reactor, a stirred autoclave, which is maintained at 80–120°C and 100 bar of ethene (Fig. 45).

b. Separation. The product alkenes are insoluble in the alcohol and phase separation takes place. After settling, the alcohol layer goes to a regeneration unit. The alkene layer is washed and ethene is recycled to the reactor. The products are distilled and the desired fractions are collected.

c. Purification and metathesis. The lower alkenes and the heavy alkenes must be "disproportionated" to give the full range of alkenes. To this end the light and heavy alkenes are sent to an isomerization reactor after having passed a purification bed, a simple absorbent to remove alcohol and ligand impurities. The isomerized mixture is then passed over a commercial molybdenum metathesis catalyst (CoMox), also a fixed bed reactor, to give a broad mixture of internal alkenes. After distillation the C_{11-14}

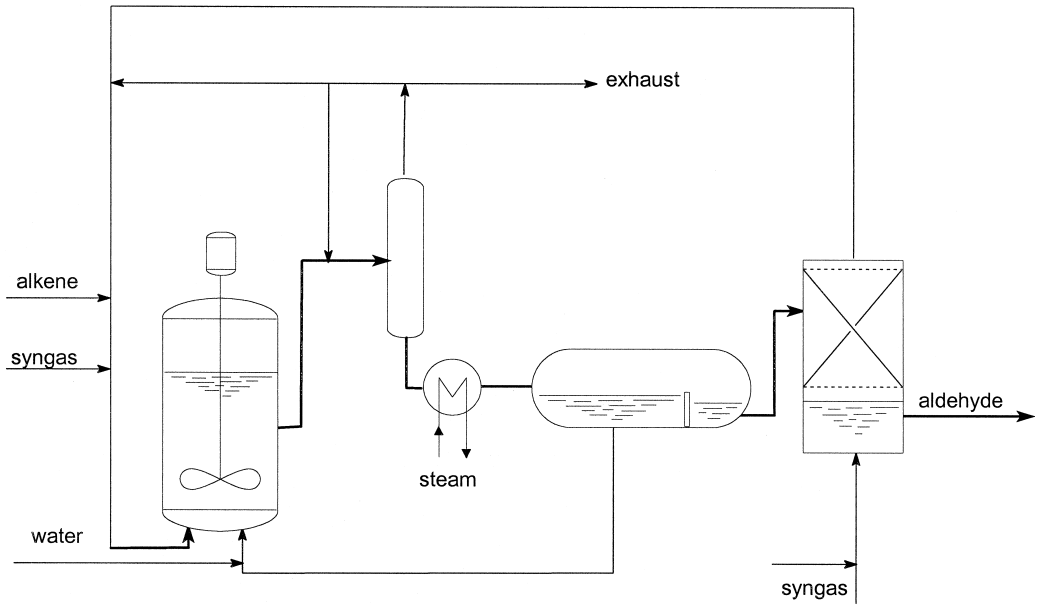


FIGURE 43 Flow-scheme of Ruhrchemie/Rhone-Poulenc process.

fraction (±15%) is used as a feedstock for alcohol production via cobalt catalysts. The light and heavy products are recycled. A bleed stream of the heavy ends must limit the buildup of branched products and polymers.

The total production of higher olefins via this and similar routes is estimated to be 2 million tons annually. A large part of the alkenes are produced for captive use.

3. Aluminum Process

Two aluminum-based processes are being used. The one-step process is operated at high temperature using aluminum alkyls as the catalyst. This process resembles the nickel process discussed; as a matter of fact Ziegler and coworkers were working with this aluminum catalyst when they accidentally found the influence of nickel, which under these drastic conditions gave the unwanted butenes.

The second aluminum-based process is a two-step process that employs aluminum stoichiometrically by each pass, the alkyls are grown at a lower temperature, decomposed in the next reactor at higher temperature, and the aluminum complexes are recycled. This procedure leads to "peaking" in the C_{10}–C_{16} range. The oligomers are now formed according to a Poisson distribution, which is very narrow compared with the Schulz-Flory distribution. When each initiator makes only one oligomer the M_w/M_n approaches 1 (Poisson) whereas M_w/M_n approaches 2 when chain transfer occurs (Schulz-Flory). In polymer synthesis these distributions also play an important role.

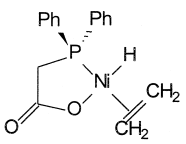


FIGURE 44 SHOP catalyst.

In the late 1990s a new group of catalysts was discovered comprising iron and cobalt complexes containing pyridinediimine ligands. Extremely fast catalysts were reported (Fig. 46). Turnover frequencies as high as several millions per hour were recorded. This represented a totally unexpected development showing that a new combination of ligand and metal can lead to surprises. Catalyst activities are very high and a separation of catalyst and product may not be needed.

F. Zirconium-Catalyzed Polymerization of Alkenes

1. Synthesis

Polymers can be made by condensation reactions and by addition polymerizations. An example of the former is the condensation of diols and diesters forming polyesters; all molecules are involved in the steady growth of species of higher molecular weight. Addition polymerization involves the reaction of initiating species with monomers, thus building up this limited number of polymer molecules

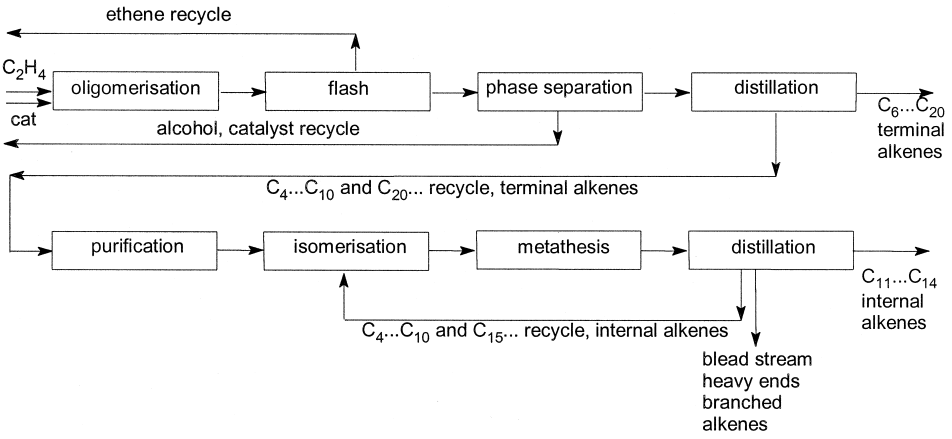


FIGURE 45 Simplified flow-scheme of SHOP.

in an excess of monomers. Four types of reaction can be distinguished for organic monomers:

1. radical polymerization;
2. anionic polymerization;
3. coordination polymerization (e.g., Ziegler-Natta polymerization); and
4. cationic polymerization.

Transition metal complexes play the key role in coordination polymerization.

2. Introduction to Ziegler-Natta Polymerization

The titanium catalyst for the stereoselective polymerization of propene to isotactic polymer was discovered by Ziegler and Natta in the mid-1950s. Soon after its discovery it was commercially exploited and even today isotactic polypropylene is one of the most important commodity polymers. Polymers with side-arms, polypropylene (PP) being the simplest one containing methyl groups, may possess stereoregularity in the arrangement of these side-arms. For example, if all methyl groups in PP are oriented toward the same direction in the stretched polymer chain, we say this polypropene is isotactic (see Fig. 47). If the methyl groups are alternatingly facing to the front and the back, we say the polymer has a syndiotactic microstructure. Random organization of the methyl groups (not shown) is found in atactic polymers. The stereoregularity of a polymer has an enormous influence on the properties of the polymer. Isotactic PP is the well-known semicrystalline material that we all know, while atactic PP is a rubber-like material. The degree of isotacticity determines the "melting point" of the polymer. Isotactic PP with less than one mistake in any of its 200 insertions melts at 166°C.

The catalyst and the concomitant technology have undergone drastic changes over the years. The titanium catalysts can be prepared by the interaction of $TiCl_4$ and alkylaluminum compounds in a hydrocarbon solvent. This reaction can be carried out with numerous variations to give a broad range of catalysts. It is a heterogeneous high-surface $TiCl_3$ material of which the active site is titanium in an unknown valence state. It is quite likely that alkyltitanium groups at the surface are responsible for the coordination polymerization. Since the 1980s, titanium catalysts supported on magnesium salts have been used.

3. The Cossee-Arlman Mechanism

The mechanism proposed for the solid titanium chloride catalysts is essentially the same for all catalysts and it is usually referred to as the Cossee-Arlman mechanism. Titanium is hexa-coordinated in the $TiCl_3$ catalysts or supported analogs by three bridging chlorides, one terminal chloride, and one terminal chloride that is replaced by an alkyl group by the alkylating agent (Et_2AlCl or Et_3Al), and a vacancy that is available for propene coordination

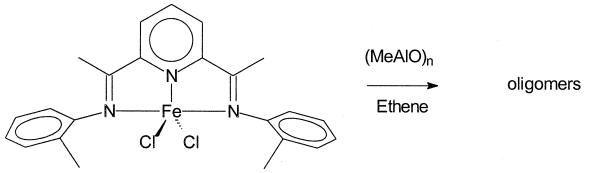


FIGURE 46 Example of new catalyst for ethene oligomerization.

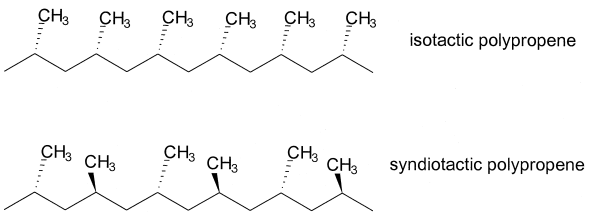


FIGURE 47 Microstructures of polypropylene.

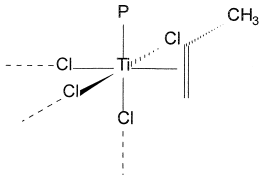


FIGURE 48 Cossee-Arlman site.

(see Fig. 48). This simple picture yields an asymmetric titanium site, but by considering the lattice surface four bridging chlorides may also lead to an asymmetric site.

In Cossee's view, one way or another, the site has to control the way a propene molecule inserts, by doing this in a very controlled manner one can imagine that a stereoregular polymer will form. This seems obvious, since different catalysts give different stereospecifities (reality is more complicated as we will see). The asymmetry of the site regulates the mode of coordination of the propene molecule, in other words it steers the direction to which the methyl group will point.

4. Site Control Versus Chain-end Control

Over the years two mechanisms have been put forward as being responsible for the stereo-control of the growing polymer chain, first the site-control mechanism and secondly the chain-end control mechanism. In the site-control mechanism the structure of the catalytic site determines the way the molecule of 1-alkene will insert (enantiomorphic-site control). As we have seen previously, propene is prochiral and a catalyst may attack either the *re*-face or the *si*-face. If the catalyst itself is chiral as the one in Fig. 48, a diastereomeric complex forms and there may be a preference for the formation of a particular diastereomer. If the catalyst adds to the same face of each subsequent propene molecule, we say isotactic PP is formed (a definition proposed by Natta). Thus, we see that stereoregular polymerization is concerned with asymmetric catalysis and indeed the way the problems are tackled these days have much in common with asymmetric hydrogenation and related processes.

When we look more closely at the intermediate polymer chain we see an alternative explanation emerging. After the first insertion has taken place a stereogenic center is obtained at carbon 2; see Fig. 49. Coordination with the next propene may take place preferentially with either the *re*-face or the *si*-face, or as displayed in Fig. 49, with the methyl group pointing up or down.

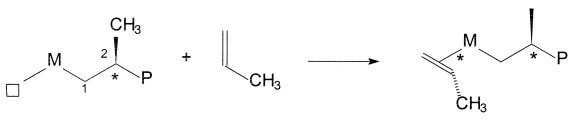


FIGURE 49 Enantiomorphic chain-end control.

In summary, in the chain-end control mechanism the last monomer inserted determines how the next molecule of 1-alkene will insert. Proof for this stems from catalysts not containing a stereogenic center that give stereoregular polymer. Secondly, whatever site-control we try to induce, the chain that we are making will always contain, by definition, an asymmetric center. These two points would strongly support the chain-end control mechanism. As we have mentioned above, the nature of the solid catalysts had an enormous influence on the product, and this underpins the Cossee site-control mechanism. Thus both are operative and both are important. Occasionally, chain-end control only suffices to ensure enantiospecifity. The *analysis* of the products using high resolution ^{13}C-NMR has greatly contributed to the mechanistic insight and distinction between the various catalysts. NMR analysis gives a detailed picture of the relative orientation of the methyl groups in the chain, i.e., the regular ones, but more in particular the mistakes that were made.

5. Homogeneous Catalysts

In the 1970s the first claims appeared concerning the homogeneous stereospecific polymerization, but they received relatively little attention as during the same years the first highly active heterogeneous titanium catalysts, immobilized on magnesium salts, were reported and the industrial interest in homogeneous catalysts diminished.

The development of the new family of homogeneous catalysts based on biscyclopentadienyl Group 4 metal complexes for the stereoselective polymerization of alkenes is mainly due to Kaminsky, Ewen, and Brintzinger. In 1980 Kaminsky and Sinn reported on an extremely fast homogeneous catalyst for the polymerization of ethene formed from the interaction of $Cp_2Zr(CH_3)_2$ and $(CH_3AlO)_n$. At 8 bar of ethene and 70°C an average rate of insertion of ethene amounting 3×10^7 mole of ethene per mole of Zr per hour was reported. For propene this catalyst led to completely atactic polymer. Ewen was the first to report the synthesis of stereoregular polymers with soluble Group 4 metal complexes and alumoxane as the co-catalyst. He found that Cp_2TiPh_2 with alumoxane and propene gives isotactic polypropene. This catalyst does not contain an asymmetric site that would be able to control the stereoregularity. A stereoblock polymer is obtained, see Fig. 50. Formation of this sequence of regular blocks is proof of the chain-end control mechanism.

Using an intrinsically chiral titanium compound (rac ethylene-bis-indenyl *titanium* dichloride), first described by Brintzinger, Ewen obtained polypropene that was in part isotactic. Kaminsky and Brintzinger have shown that

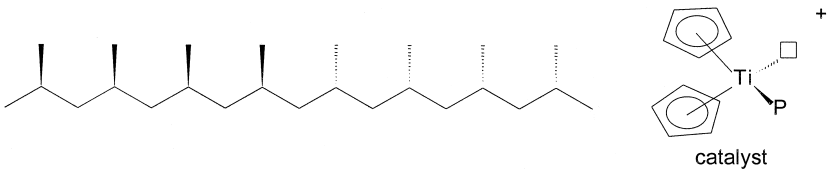


FIGURE 50 Stereo-block isotactic PP with an achiral catalyst.

highly isotactic polypropene can be obtained using the racemic *zirconium* analog of the ethylene-bis(indenyl) compound. Modification of the cyclopentadienyl ligands has led to a very rich chemistry and a great variety of microstructures and combination thereof has been obtained including isotactic polymer with melting points above 160°C, syndiotactic polypropene, block polymers, hemi-isotactic polymers, etc.

6. Mechanistic Explanations

Fundamental studies have led to a detailed insight into the mechanism of the polymerization and the control of the microstructure through the substituents on the cyclopentadienyl ligands. The nature of the catalyst has been the topic of many studies and is now generally accepted to be a cationic Cp_2ZrR^+ species. The first requirement for obtaining an active catalyst is to ascertain the formation of cationic species. Secondly, the metallocene (or related ligand) must have the correct steric properties.

The achiral Cp_2TiPh_2 catalyst activated with alumoxane can produce PP with some isotacticity. Hence, the natural tendency of the chain in this 1,2 insertion is to catalyze the formation of isotactic PP. How do we explain the improvements made with the bridged, chiral metallocenes? We need a catalysts with a specific preference for coordination to either the *re*- or *si*-face of the incoming propene. The metallocene can be equipped with a chiral substituent and thus we can try if this is indeed effective. By trial and error and molecular modeling we will find the best substituent at our Cp-ring. Two sites are needed in our catalyst: one for the migrating chain and the other for the propene molecule. We know that insertion takes place in a migratory manner. The one asymmetric site that we created controlling the coordination of propene is insufficient, because after this migration the alkyl chain will reside on the other site. The next propene molecule will coordinate to the former alkyl site. Thus, the two sites must be the same and yet chiral. In asymmetric hydrogenation we have seen how this problem can be solved, viz. with a C_2 symmetric site. The first compounds introduced having this feature are shown in Fig. 51.

Zirconium is tetrahedrally surrounded by two cyclopentadienyl ligands and two chlorides. The latter two ligands are not essential as they are replaced before the polymerization can start by an alkyl group (e.g., CH_3 from methylalumoxane) and by a solvent molecule, thus generating the required cationic species.

The two cyclopentadienyl anions are linked together by a bridge, here an ethane bridge, and extended with an organic moiety which renders the molecule its chirality. In the example of Fig. 51 a 1,2-ethane bridged bis(1,1-tetrahydroindenyl) dianion has been drawn. Two compounds are obtained when this complex is synthesized, the rac-isomer mixture and the meso-isomer. Often they can be separated by crystallization. The sterically less hindered rac-isomers may be formed in excess. The isomers can only interchange by breaking of a cyclopentadienyl-metal bond and reattaching the metal at the other face of the cyclopentadienyl ligand. This process is very slow compared to the rate of insertion reactions with propene.

The rac-isomers have a twofold axis and therefore C_2-symmetry. The meso-isomer has a mirror plane as the symmetry element and therefore C_s-symmetry. For polymerization reactions the racemic mixture can be used since the two chains produced by the two enantiomers are identical when begin- and end-groups are not considered. The two sites (X, in Fig. 51) have the same absolute configuration, and coordination at both sites will occur with a preference for the same face of the propene molecule. A simple way of looking at it would be that the cyclohexyl rings enforce a certain position of the methyl group of propene. When

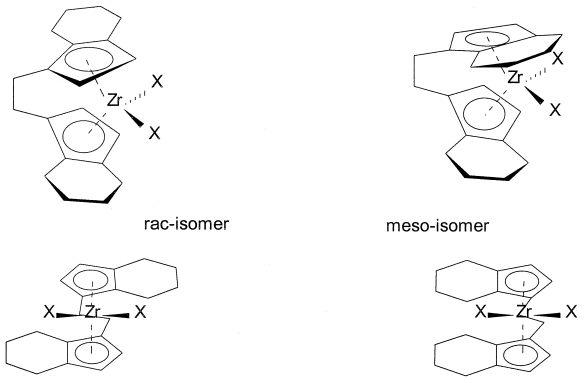


FIGURE 51 Structure of 1,2-ethanediyl-tetrahydro-indenyl-ZrX_2.

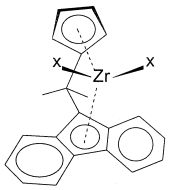


FIGURE 52 Bridged fluorenyl-cyclopenadienyl complex giving syndiotactic PP.

we rotate the molecule around the twofold axis we see that the site carrying propene is equivalent to the other. In Natta's words, an isotactic polymer should form (neglecting mistakes, and thermodynamic versus kinetic factors). The meso-isomer has no preference for either face of propene and gives atactic polypropylene.

7. Site and Chain-End Control Reenforce Isotactic Specificity

When dealing with cationic titanium and zirconium metallocene catalysts, the insertion is a 1,2 mode. As a result, the two control mechanisms reenforce one another. Indeed, the best results in terms of isotacticity are obtained with these catalysts.

8. Syndiotactic PP by Site Control

A catalyst that contains two binding sites with the same absolute configuration gives isotactic PP. When we can make a catalyst that contains two identical sites but with the opposite chirality, each site should coordinate to the opposite face of propene. Following Natta's definition this should give syndiotactic polymer. In other words, a catalyst containing a mirror plane—in such a way that the two coordination sites are mirror images—gives syndiotactic polymer, because one site should coordinate to one face of propene, and vice versa. To this end, Ewen made isopropyl(1-fluorenyl-cyclopentadienyl) ligands and their metallocenes; see Fig. 52. This complex has no chirality (i.e., the dichloride precursor). The catalyst indeed gives syndiotactic polymer.

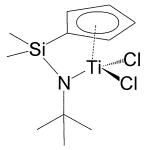


FIGURE 53 "Single site" catalyst for polyethylene.

The above initial findings were the beginning of a new generation of catalysts for making polyolefins. It has initiated a lot of research both in industry and academia. Especially interesting is the molecular design of both catalysts and polymers. For process reasons, however, the new homogeneous catalysts, for polypropylene at least, should be heterogenized on solid supports.

Only one commercialization of the new metallocene catalysts will be mentioned here as an example, the so-called "Single Site Catalyst" developed by Dow and shown in Fig. 53. It is not a stereospecific catalyst as it simply polymerizes ethene and adds higher alkenes. It gives a narrow MW and a high rate of (re-)insertion of higher alkenes. The control of branching and the extent of long-chain branching leads to a product that can be easily processed.

G. Palladium Catalysis

Palladium catalysts are nowadays an important ingredient of many multistep syntheses both in industry and the laboratory. The versatility of palladium is enormous. Here we will present a few typical examples. In several instances iso-electronic complexes of nickel and sometimes rhodium can be used or they lead to even better results. The basic principles are the same.

1. Allylic Alkylation

The palladium-catalyzed allylic alkylation has become a very important tool for organic chemists. It has been developed by Trost, following an initial report by Tsuji. The reagents used are all very mild and are compatible with many functional groups. The method has been applied in

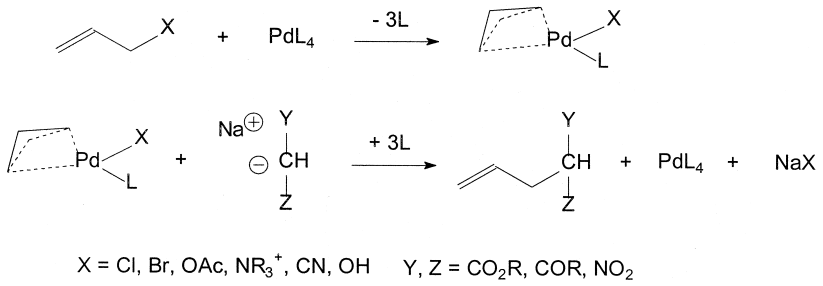


FIGURE 54 Palladium(0)-catalyzed allylic substitution.

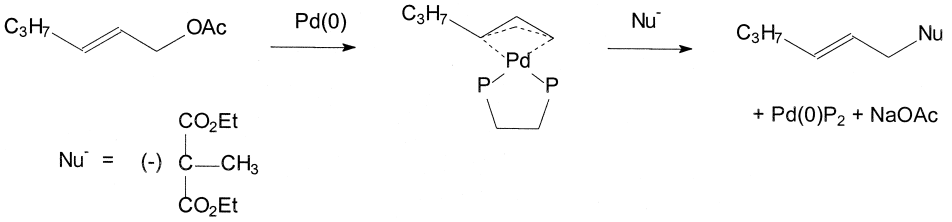


FIGURE 55 Regioselectivity in allylic substitution.

the synthesis of many complex organic molecules. During the reaction a new carbon-carbon bond is formed and the resulting molecule still contains a double bond that might be used for further derivatization.

The reaction starts with an oxidative addition of an allylic compound to palladium zero. Allyl halides, carboxylates, etc., can be used. At first we will consider triphenylphosphine as the ligand, but often ***large ligand influences have been detected. A π-allyl-palladium complex forms. Formally, the allyl group is an anion in this complex, but owing to the high electrophilicity of palladium, the allyl group undergoes attack by nucleophilic reagents, especially soft nucleophiles. After this attack, palladium(0) "leaves" the allyl group and the product is obtained. Palladium zero can reenter the cycle and hence the reaction is catalytic in palladium. As a side-product a salt is formed. For small-scale industrial applications it is not a problem to make salts in stoichiometric amounts. The mechanism and a few reagents are shown in Fig. 54.

Regioselectivity is high in this reaction and depends on the ligand used. Ligand effects can ensure substitution at the allylic carbon carrying most carbon substituents, or just the reverse. When a diphosphine ligand is used a hexenyl group instead of the allyl group substitution occurs mainly at the terminal carbon, see Fig. 55.

An asymmetric application of this reaction has been developed. The model substrate studied most is the symmetrically 1,3-diphenyl substituted allylpalladium complex shown in Fig. 56. The ligand used is a C_2 chiral diphosphine ligand, but also C_1 asymmetric ligands have been successfully applied. The orientation of the two phenyl substituents at carbons 1 and 3 of the allyl fragment is different under the influence of the chiral C_2 ligand. Without the chiral C_2 ligand carbons 1 and 3 are mirror images; palladium is attached to the "local" *re*- and *si*-face. The chiral C_2 ligand makes carbons 1 and 3 diastereotopic, i.e., they are chemically different. As a result one specific carbon atom (carbon 1) undergoes selective attack by the nucleophile. This way a chiral compound is obtained.

A great variety of ligands have been used. Trost has been especially successful using a chiral bidentate phosphine with a very large bite angle (110°); see Fig. 57.

This ligand "embraces" the metal and thus exerts its influence on the allyl group. Even substrates carrying small substituents can now be asymmetrically substituted.

2. Cross Coupling

The making of carbon-to-carbon bonds from carbocations and carbo-anions is a straightforward and simple reaction. Easily accessible carbo-anions are Grignard reagents RMgBr and lithium reagents RLi. They can be conveniently obtained from the halides RBr or RCl and the

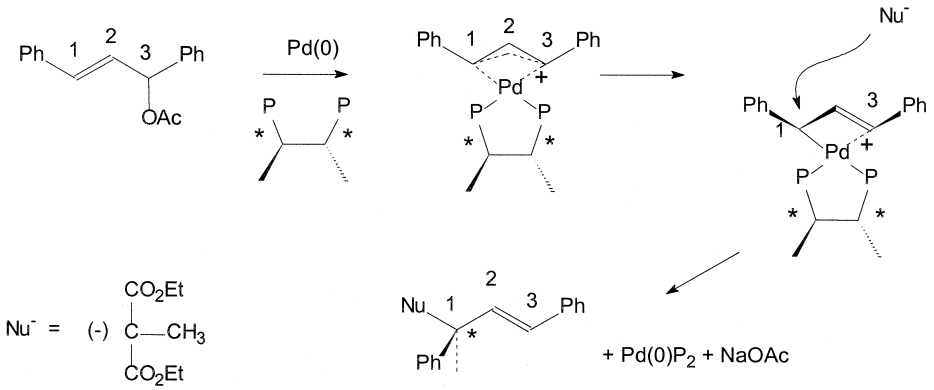


FIGURE 56 Asymmetric allylic substitution.

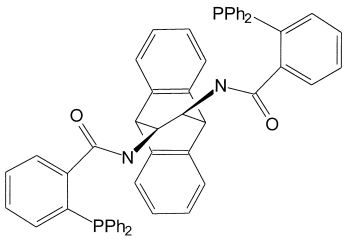


FIGURE 57 One of Trost's asymmetric ligands for "AAA," asymmetric allylic alkylation.

metals Mg and Li. They are both highly reactive materials, for instance, with respect to water.

The reactions of these carbon centered anions with polar compounds such as esters, ketones, and metal chlorides are indeed very specific and give high yields. The reaction of Grignard reagents and the like with alkyl or aryl halogenides, however, is extremely slow giving many side-products, if anything happens at all.

The "cross coupling" reaction has found wide application both in organic synthesis in the lab and in industrial environment. The transition metal catalysts are usually nickel and palladium. In addition to organomagnesium and organolithium a great variety of organometallic precursors can be used. Also, many precursors can serve as starting materials for the carbocation. Last but not least, the ligand on the transition metal plays an important role in determining the rate and selectivity of the reaction. Here we will present only the main scheme and take palladium as the catalyst example. Figure 58 gives the general scheme of the palladium-catalyzed cross-coupling reaction.

We can start with palladium(II) or palladium(0), but for the present explanation the latter is more convenient. Oxidative addition of an aryl halide (PhBr in Fig. 58) to palladium zero takes place and a square planar Pd(II) complex is formed. Subsequently the inorganic bromide reacts with the Grignard or lithium alkyl reagent (here 2-BuMgBr) giving a diorganopalladium complex and magnesium dibromide. The third reaction that occurs is the reductive elimination giving the organic cross-coupling product and the palladium(0) catalyst in its initial state.

Less reactive organometallics derived from tin and boron can also be used. These reagents do not react with water, but they are still able to alkylate the palladium bromide intermediate. As mentioned above, their formation does involve one more step because they are also made via Grignard type reagents. The coupling using tin organometallics is referred to as "Stille" couplings. The second reagent based on boron was introduced by Suzuki. The boronic ester derivative is made from trimethyl borate and an aryl anion reagent followed by hydrolysis of the two remaining methyl ester groups. This phenylboronic acid is soluble in water and the coupling reaction can even be carried out in water. See Fig. 59 for a scheme of the Suzuki coupling.

The cross-coupling reaction is applied industrially in the synthesis of alkyl-aryl compounds that are used in liquid crystals, aryl-aryl compounds in agrochemicals, pharmaceuticals, etc.

3. Heck Reaction

A reaction related to the cross coupling is the Heck reaction. The reaction involves the coupling of an aryl halide (or pseudo halide) with an alkene in the presence of a base. Owing to the latter, as in the cross-coupling reaction an equivalent of base is formed. In the base case, the Heck reaction will produce substituted styrenes. These products can also be made in the cross-coupling reaction discussed above. It is attractive that the Heck reaction does not involve Grignard-type reagents and thus it allows the presence of groups reactive toward Grignard reagents such as esters, acids, ketones, etc.

The reaction starts again with the oxidative addition of an aryl halide (Br or I) to palladium zero. The next step is the insertion of an alkene into the palladium carbon

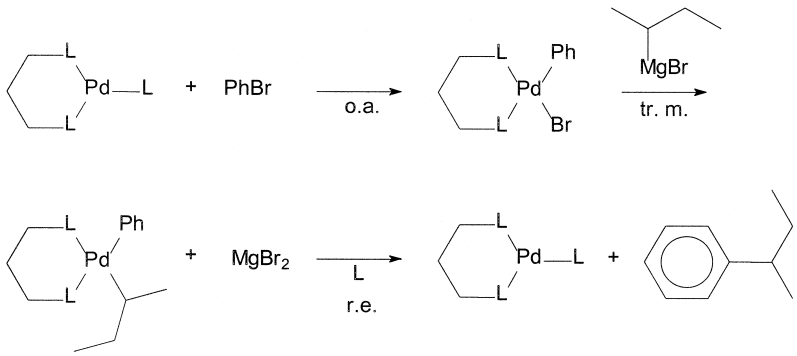


FIGURE 58 Palladium-catalyzed cross coupling.

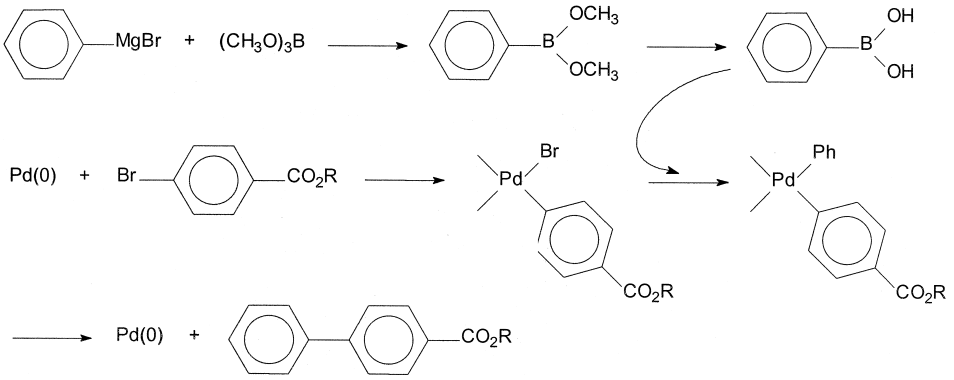


FIGURE 59 Suzuki cross-coupling reaction.

bond just formed. The third step is β-elimination giving the organic product and a palladium hydrido halide. The latter reductively eliminates HX that reacts with base to give a salt. Figure 60 shows the reaction scheme.

In modern Heck chemistry the halide is replaced by noncoordinating anions such as O_3SCF_3. The advantage is that the palladium center is now much less saturated and more positively charged (and hence more reactive) during the alkene coordination and insertion steps.

4. Wacker-type Reactions

The Wacker process is the oxidation of ethene by divalent palladium to ethanal in the presence of water. The Wacker-Hoechst process has been studied in great detail and in all textbooks it occurs as *the* example of a homogeneous catalyst system illustrating nucleophilic addition to alkenes. Palladium chloride is the catalyst and palladium activates coordinating ethene toward a nucleophilic attack by water. The overall reaction reads:

$$C_2H_4 + 1/2\ O_2 \rightarrow CH_3CHO \qquad \Delta H = -244\ \text{kJ/mol}$$

Thus the reaction is highly exothermic as one might expect for an oxidation reaction.

Figure 61 shows the reaction scheme. First coordination of ethene to palladium has to take place. One chloride ion is replaced by water and one by ethene. Ethene is activated toward nucleophilic attack by coordination to the electrophilic palladium ion. Then the key step occurs, the attack of water (hydroxide) to the activated ethene molecule. The nucleophilic attack of water or hydroxide takes place in an anti-fashion; i.e., the oxygen attacks from outside the palladium complex and the reaction is not an insertion of ethene into the palladium oxygen

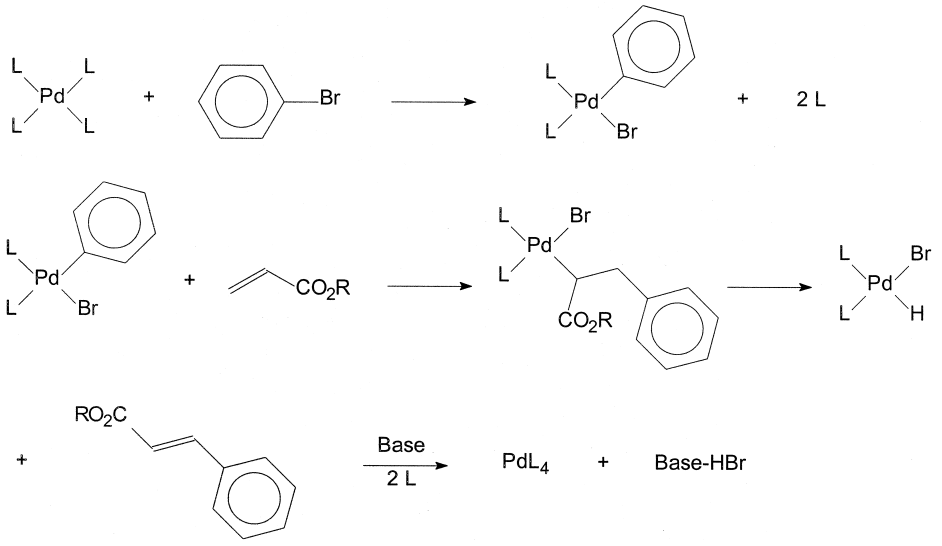


FIGURE 60 Heck reaction.

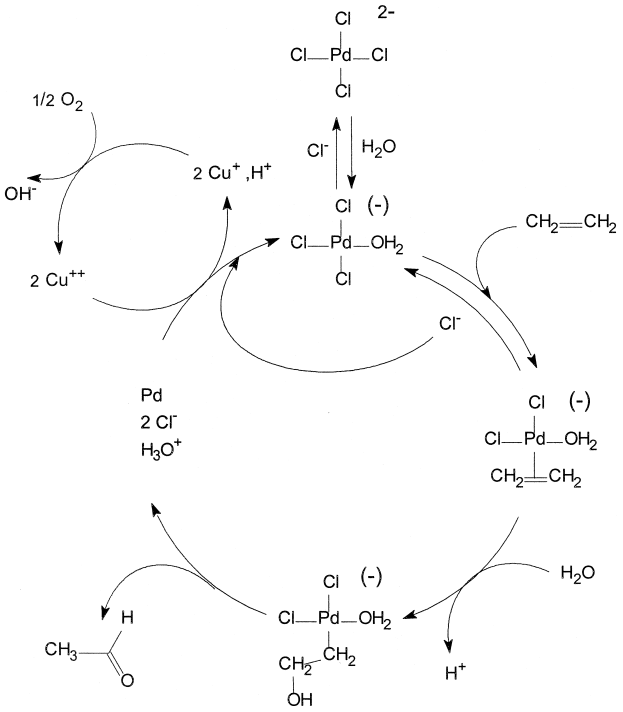


FIGURE 61 Reaction scheme of the Wacker ethanal production.

bond. The nucleophilic attack of a nucleophile on an alkene coordinated to palladium is typical of "Wacker-" type reactions. The rate equation is in agreement with the replacement reactions:

$$v = k\left[\mathrm{PdCl_4^{2-}}\right][\mathrm{C_2H_4}]\left[\mathrm{H_3O^+}\right]^{-1}[\mathrm{Cl^-}]^{-2}$$

After a rearrangement of the hydroxyethyl group, acetaldehyde (ethanal) forms and palladium zero is the reduced counterpart. Including the palladium component the reaction reads:

$$\mathrm{C_2H_4 + Pd^{2+} + H_2O \rightarrow Pd(0) + 2H^+ + CH_3CHO}$$

This reaction had been known since the beginning of the century. Reoxidation of palladium directly by oxygen is extremely slow. The invention of Smidt (Wacker Chemie) involved copper as the intermediate in the recycle of divalent palladium:

$$\mathrm{2\ Cu^{2+} + Pd(0) \rightarrow 2Cu^+ + Pd^{2+}}$$

$$\mathrm{2\ Cu^+ + 1/2\ O_2 + 2H^+ \rightarrow 2\ Cu^{2+} + H_2O}$$

All kinetics are in support of the formation of a β-hydroxyethylpalladium complex as the intermediate.

An important application of the Wacker chemistry is the reaction of ethene, acetic acid, and dioxygen over a heterogeneous catalyst containing palladium and a reoxidation catalyst for the commercial production of vinyl acetate.

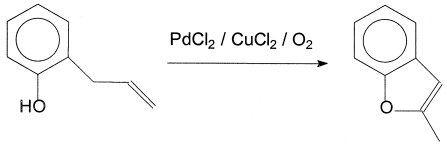


FIGURE 62 Oxypalladation of 2-allylphenol.

5. Wacker-type Reactions

A recent development is the use of Wacker activation for intramolecular attack of nucleophiles to alkenes in the synthesis of organic molecules. The nucleophilic attack is intramolecular; the rates of intermolecular reactions are very low. The reaction has been applied in a large variety of organic syntheses and is usually referred to as Wacker (type) activation of alkene (or alkynes). If oxygen is the nucleophile, we also say oxypalladation. An example is shown in Fig. 62. Under these conditions the palladium catalyst is often also a good isomerization catalyst, which leads to the formation of several isomers.

As shown above, palladium can catalyze a variety of reactions. In addition to the ones shown palladium can also catalyze insertions of carbon monoxide, followed by ester formation with alcohols or aldehyde formation with dihydrogen. Combinations of these reactions in one system are known. From a point of view of economy this can be very attractive as several steps are combined in one reactor, using just one catalyst. These "cascade-" or "tandem-" type reactions will become very important for fine chemical synthesis in the near future.

H. Asymmetric Epoxidation

1. Introduction

Epoxidation of alkenes (Fig. 63) is a powerful reaction for the introduction of oxygen in hydrocarbons. Oxygen can be used for the oxidation of ethene over a silver catalyst, but for all other alkenes hydroperoxides and related reagents are the source of oxygen to avoid further oxidation. As catalysts one can use titanium alkoxides, and molybdenum and tungsten compounds.

Epoxides are interesting starting materials for further derivatization. In most instances the reaction will lead to the formation of mixtures of enantiomers (that is when the

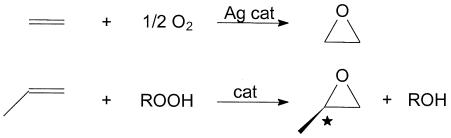


FIGURE 63 Catalytic epoxidation of alkenes.

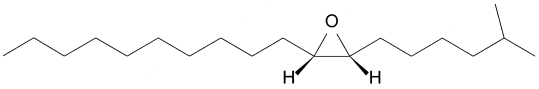


FIGURE 64 Disparlure.

alkenes are prochiral, or when the faces are enantiotopic). Three important reactions should be mentioned in this context the:

1. Katsuki-Sharpless epoxidation of allylic alcohols
2. Sharpless asymmetric hydroxylation of alkenes with osmium tetroxide
3. Jacobsen asymmetric epoxidation of alkenes

All three reactions find wide application in organic synthesis. Here we will only discuss the Sharpless epoxidation of allylic alcohols. This reaction finds industrial application in Arco's synthesis of glycidol, the epoxidation product of allyl alcohol, and Upjohn's synthesis of disparlure (Fig. 64), a sex pheromone for the gypsy moth. The synthesis of disparlure starts with a C_{10} allyl alcohol in which the alcohol is replaced by the other carbon chain after asymmetric epoxidation. Perhaps today the Jacobsen method can be used directly on a suitable alkene, although the steric differences between both ends of the molecules are extremely small.

2. Katsuki-Sharpless Asymmetric Epoxidation

The reaction uses allylic alcohols and hydroperoxides. The catalyst is a chiral Ti(IV) catalyst. The chirality is introduced in the catalyst by reacting titaniumtetraisopropoxide with one mole of a simple tartrate diester. The reaction is shown in Fig. 65.

The enantioselectivity is not very sensitive to the nature of the allylic alcohol. By contrast, titanium and tartrates are essential to the success. Note the difference with the L-dopa asymmetric hydrogenation, which can be carried out with hundreds of C_2 chiral diphosphines and specific substrates only.

I. Terephthalic Acid

1. Introduction

Terephthalic acid (1,4-benzenedicarboxylic acid) is used for the production of polyesters with aliphatic diols as the comonomer. The polymer is a high-melting, crystalline material forming very strong fibers. It is the largest volume synthetic fiber and the production of terephthalic acid is the largest scale operated process based on a homogeneous catalyst. More recently the packaging applications (PET, the recyclable copolymer with ethylene glycol) have also gained importance. Terephthalic acid is produced from *p*-xylene by oxidation with oxygen. The reaction is carried out in acetic acid and the catalyst used is cobalt (or manganese) acetate and bromide. Phthalic anhydride is made from naphthalene or *o*-xylene by air oxidation over a heterogeneous catalyst. The main application of phthalic anhydride is in the dialkylesters used as plasticizers (softeners) in PVC. The alcohols used are, for instance, 2-ethylhexanol obtained from butanal, a hydroformylation product.

2. The Chemistry

The overall reaction reads as follows (Fig. 66):

The reaction produces two moles of water per mole of terephthalic acid. The oxidation reaction is a radical process. The key step of the reaction scheme involves the cobalt(III)-bromide-catalyzed H-abstraction to give a benzylic radical, hydrogen bromide, and a divalent cobalt species (Fig. 67). This is the initiation reaction of the radical chain reaction. Without cobalt or manganese and bromide the reaction would be very slow.

When one methyl group has been oxidized 4-toluic acid is formed. This intermediate is soluble in the solvent used, acetic acid. In the next step it is oxidized to terephthalic

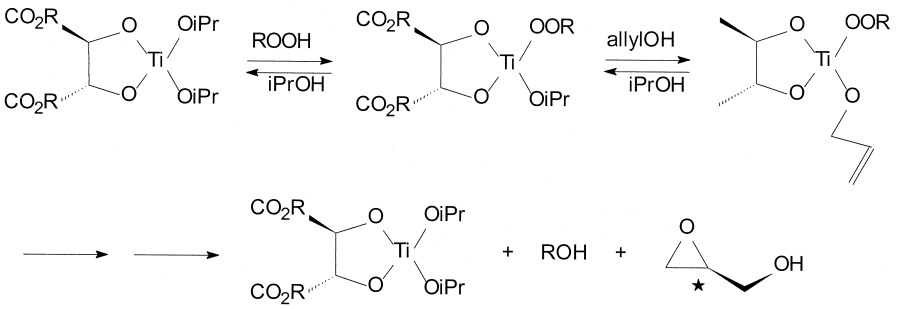


FIGURE 65 Katsuki-Sharpless asymmetric epoxidation.

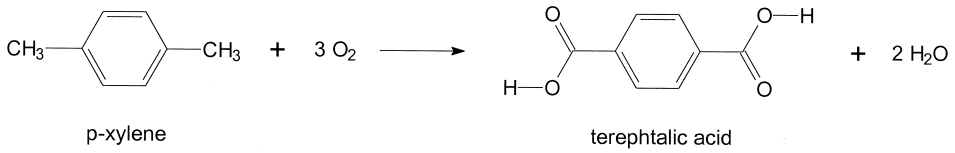


FIGURE 66 Formation of terephthalic acid.

acid, which is almost insoluble in acetic acid. The intermediate product is 4-formylbenzoic acid. This cocrystallizes with the final product. Since it is a mono-acid it is will cause a termination of the polymerizing chain.

High molecular weights are needed to obtain strong materials. Thus, the presence of mono-acids is detrimental to the quality of terephthalic acid and they have to be carefully removed. One approach involves a second oxidation carried out at higher temperatures and the other method involves reduction of the formylbenzoic acid to toluic acid. The latter in more soluble and stays behind in the liquid. The reduction is done with a palladium catalyst on carbon support. This method gives the highest quality terephthalic acid.

The polymerization with ethylene glycol can be carried out directly from the acid if this is very pure. The esterification/polymerization is conducted at high temperatures (200–280°C) using metal carboxylates as homogeneous catalysts.

3. Process Description

Most of the terephthalic acid is produced with a catalyst system developed by Scientific Design. It was purchased by Amoco and Mitsui and is referred to as the Amoco Oxidation. The solvent is acetic acid. Compressed air is used as the source for oxygen. The catalyst dissolved in acetic acid and the two reactants are continuously fed into the reactor. The temperature is around 200°C and the pressure is approximately 25 bar. The reaction is very exothermic (1280 kJ/mol). This heat of reaction is removed by evaporation/condensation of acetic acid. This heat can be used in the solvent distillation/purification part of the plant; see Fig. 68.

The oxidation proceeds in two steps, *p*-toluic acid is the intermediate. This mono-acid has a high solubility in this medium and undergoes the second oxidation. The conversion is high, approaching 100% and the yield based on *p*-xylene is ~95%. The solubility of terephthalic acid in most solvents is extremely low, even at this high temperature. The solubility in acetic acid at room temperature is 0.1 g/l. At 200°C it is still only 15 g/l. Thus, most of the product crystallizes while it is being formed and, as mentioned above, the intermediate formylbenzoic acid cocrystallizes. The mixture of liquid and solid is led to a filter in which the crude product containing 0.5% of mono-acid is collected. The liquid, containing traces of xylene, *p*-toluic acid, water, the catalyst, and some product is sent to a distillation unit to remove water and to recover acetic acid and the catalyst. The catalyst is relatively cheap, but the heavy metals cannot be simply discarded and a recycle of these robust catalysts would seem to be in place.

As outlined above a purity of "only" 99.5% is a serious problem in making a polymer feedstock and the mono-acid intermediate has to be removed either by oxidation under more forcing conditions (Mitsubishi) or hydrogenation to the more soluble 4-toluic acid (Amoco). The final product contains amounts of 4-formylbenzoic acid as low as 15 ppm. The proportion of toluic acid may be an order of magnitude higher. Aldehydes usually cause coloring of polymers and that is an additional reason to remove them carefully.

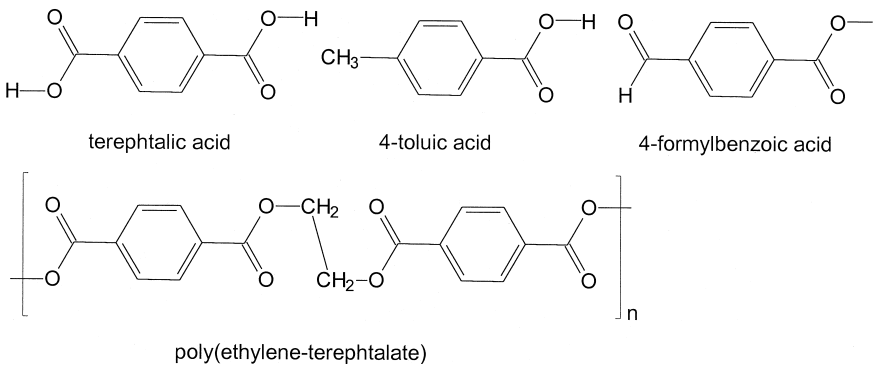


FIGURE 67 Structures involved in terephthalic acid synthesis.

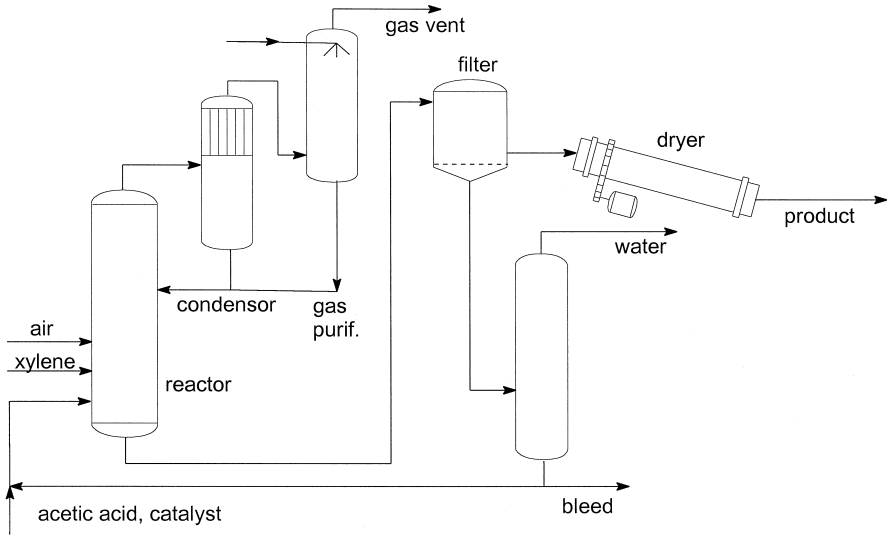


FIGURE 68 Simplified scheme for *p*-xylene oxidation.

J. Adiponitrile by Hydrocyanation

The transition metal-catalyzed addition of HCN to alkenes is potentially a very useful reaction in organic synthesis, and it certainly would have been widely applied if its attraction was not largely offset by the toxicity of HCN. Industrially the difficulties can sometimes be minimized to an acceptable level, and Du Pont has commercialized the addition of HCN to butadiene for the production of adiponitrile ($NC(CH_2)_4CN$), a precursor to 6,6-nylon.

The catalyst precursor is a nickel(0) tetrakis (phosphite) complex which is protonated to form a nickel(II) hydride. Alternatively, we could write an oxidative addition of HCN to nickel zero. Subsequently coordination and insertion of the diene takes place followed by reductive elimination of a pentenenitrile with concurrent regeneration of the nickel zero complex. Two isomerization reactions must occur in order to achieve a high selectivity, and then the HCN addition cycle is performed for the second time on the much less reactive 4-pentenenitrile. In Fig. 69 the hydrocyanation mechanism in a simplified form is given for ethene; the basic steps are the same as for butadiene but the complications due to isomer formation are lacking.

Hydrocyanation of butadiene is more complicated than that of ethene; it requires two hydrocyanation steps and several isomers can be observed. The isomers obtained in the first step of the HCN addition to butadiene are shown in Fig. 70. The addition first leads to compounds **1** and **2**, which equilibrate perhaps via the retroreaction. In order to obtain adiponitrile, **2** should isomerize to **4**, and not to the thermodynamically more stable **3** (conjugation). The thermodynamic ratio is **2**:**3**:**4** = 20:78:1.6. The isomerization of **2** to **4** happens to be favorably controlled by the kinetics of the reactions; the reaction **2** to **4** reaches equilibrium, but the reaction **2** to **3** does not. Note that the nickel complex not only is responsible for the addition of HCN but that it is also capable of catalyzing selectively the isomerization. The final step is the addition of HCN to **4** to give **5**, adiponitrile.

K. Alkene Metathesis

Metathesis of alkenes has been known for quite some time. The initial catalysts were based on tungsten, molybdenum, and rhenium. Both homogeneous and heterogeneous catalysts find application. Well-known, older homogeneous catalysts can be formed from WCl_6 and an alkylating reagent. After dialkylation with reagents such as $EtAlCl_2$

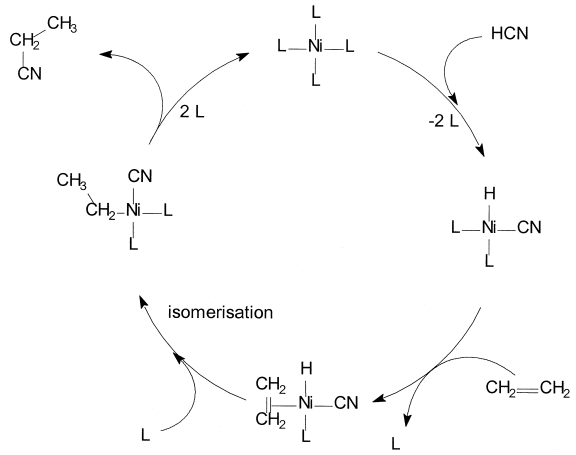


FIGURE 69 Hydrocyanation of ethene.

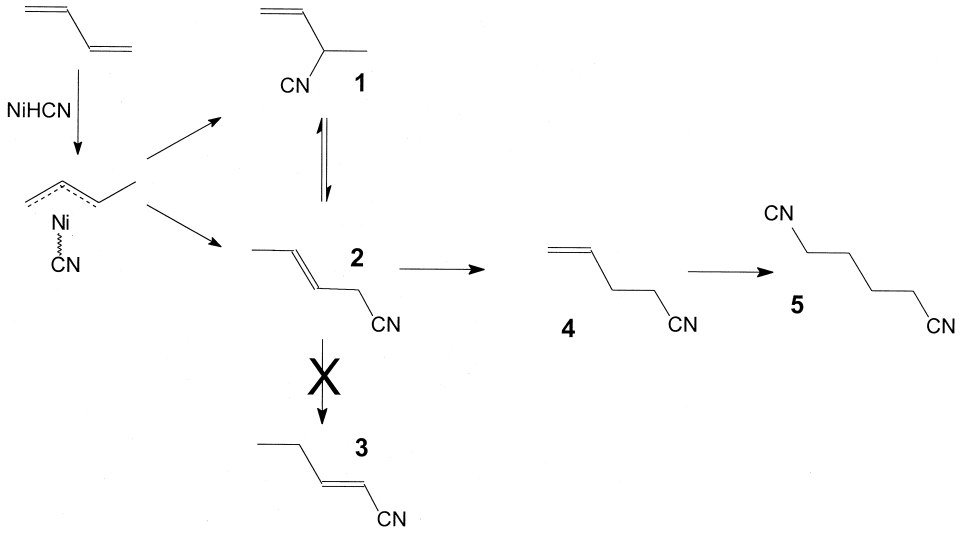


FIGURE 70 Hydrocyanation of butadiene leading to adiponitrile.

or R_4Sn, α-elimination occurs and a metal alkylidene is formed, $Cl_4W{=}CR_2$, see elementary steps. Often oxygen containing promoters including O_2, EtOH, or PhOH are added. Turnover frequencies as high as 300,000 mol (product) $\cdot$ mol^{-1}(cat) $\cdot$ h^{-1} can be achieved, even at room temperature. The development of *well-defined catalysts* based on molybdenum and tungsten is mainly due to Schrock. Synthetically useful reactions, including acyclic olefin metathesis, ring-opening metathesis polymerization (ROMP), alkyne polymerization, acyclic diene metathesis polymerization (ADMET), and ring-closing metathesis (RCM) have been catalyzed by early-transition-metal alkylidenes. Porri introduced *in situ* catalysts based on ruthenium or iridium in polar media. The synthesis of well-defined ruthenium alkylidene complexes of the type $RuCl_2(=CHPh)(PR_3)_2$ was developed by Grubbs; see Fig. 71.

The peculiar property of the ruthenium (and also iridium catalysts) is that they are active in the presence of polar substituents. The initial catalysts, generated from high-valence metal chlorides with alkylating agents were not resistant to oxygen-containing compounds. Schrock developed sterically hindered alkylidene complexes, such as the one shown in Fig. 72. The *in situ* prepared catalysts containing bulky phenols have also been reported. By replacing the two alkoxides by a chiral bulky bisnaphthol asymmetric metathesis has been achieved.

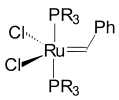


FIGURE 71 An example of Grubbs' ruthenium catalysts.

Many applications of these new catalysts to organic chemistry are known as they can serve as a means to make large rings, for instance, after two allyl groups have been introduced to a nonlinear fragment; see Fig. 73. The resulting double bond can be further functionalized or hydrogenated if needed.

L. Cyclopropanation

As in the previous section about metathesis, cyclopropanation is concerned with the transfer of carbene or alkylidene species from a metal to an organic molecule. The reaction involves the metal-catalyzed decomposition of a diazo compound to give a reactive metal-alkylidene complex followed by the transfer of the "carbene" to an alkene. The product of this very convenient reaction is a cyclopropane derivative in a single step (see Fig. 74). Catalysts used for this reaction are based on copper, rhodium, and ruthenium. A common, relatively stable diazo compound is N_2CHCO_2Et. The example shown in Fig. 74 is carried out commercially using a copper complex containing a chiral imine dialkoxide ligand. After conversion of the

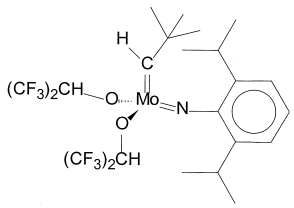


FIGURE 72 Schrock's catalyst for metathesis of functional alkenes.

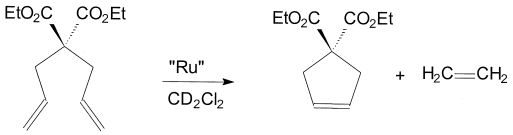


FIGURE 73 Ring-closing metathesis.

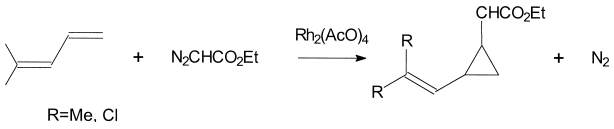


FIGURE 75 Cyclopropanation yielding pyrethroid insecticides.

acid to a substituted amide Cilastatin, a pharmaceutical is obtained.

Cyclopropanation of dienes leads to an important class of compound used as insecticides, the pyrethroid-type compounds related to chrysanthemic acid, a natural insecticide obtained from East African chrysanthemums. Catalysts used comprise rhodium (II) dimers of carboxylates and related compounds. Chiral complexes can lead to high enantioselectivities. This chemistry has been successfully developed by Doyle. In Fig. 75 the synthesis of the basic skeleton has been depicted, which requires replacement of the ethyl group by other groups. Interestingly, the "carbene" fragment can also be inserted into N−H or C−H bonds. The former reaction is used for making β-lactams.

The activity of ruthenium complexes for this reaction affords a bridge between the metathesis reaction in the previous section and cyclopropanation, an area that has been explored by Noels. A common intermediate can be imagined for the two reactions leading either to metathesis or cyclopropanation, as shown schematically in Fig. 76.

M. Hydrosilylation

The catalytic addition of a silicon compound containing an Si−H bond to a C=C double bond is the most common way for the synthesis of silane or silicon compounds. The silicon compounds are used to make polymers (silicone rubbers) or a large variety of agents used for the modification of surfaces. For making polymers the reagent used is, for instance, methyldichlorosilane, which is added to vinyl compounds using a catalyst (Fig. 77). The dichlorosilane compounds are hydrolyzed to give silicone polymers.

For industrial processes the oldest and preferred catalyst for the hydrosilylation reaction is Speier's catalyst, H_2PtCl_6, which can be used in parts per million quantities. The catalyst is reduced *in situ* and probably alcohols and/or other oxygen-containing compounds are involved in this reaction. Schematically the reaction mechanism can be imagined as a sequence involving alkene complexation, oxidative addition of the silane breaking the Si−H bond, and insertion and reductive elimination, as shown in Fig. 78. It is referred to as the Chalk-Harrod mechanism and it follows exactly the elementary steps described in the introduction. The precise nature of the platinum catalyst for this process, however, has not been delineated. The catalysts are extremely sensitive to the conditions applied and in practice one encounters incubation times, capricious kinetics, alkene isomerization, formation of unsaturated side products, etc.

A large variety of metal complexes catalyze the hydrosilylation reaction. Other double bonds can serve as the silane acceptor, such as alkynes, ketones, and imines. For the latter two substrates asymmetric catalysts also have been developed.

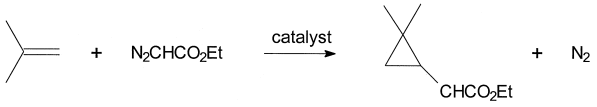


FIGURE 74 Cyclopropanation using diazocompounds as "carbene" precursor.

N. Polyesters

One of the largest applications of homogeneous catalysis involves the manufacture of polyesters. In a simple condensation process diacids and diols are polymerized forming the polymers. The simplest one is polyethyleneterephthalate (PET), which is made from ethylene glycol and terephthalic acid as discussed above.

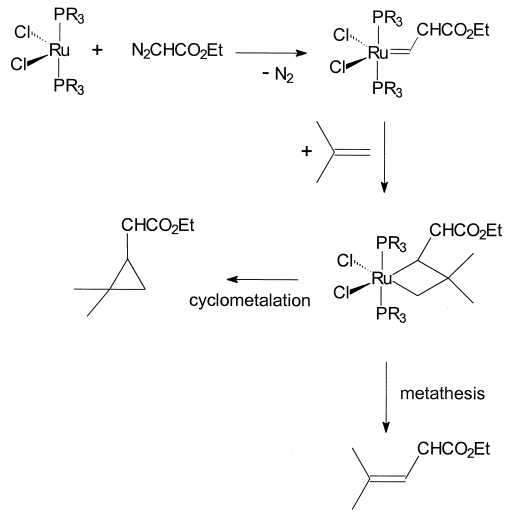


FIGURE 76 Relation between metathesis and cyclopropanation mechanism.

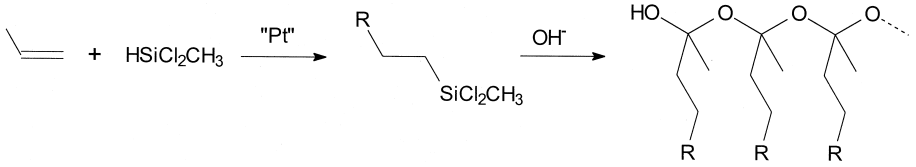


FIGURE 77 Hydrosilylation and silicone polymer formation.

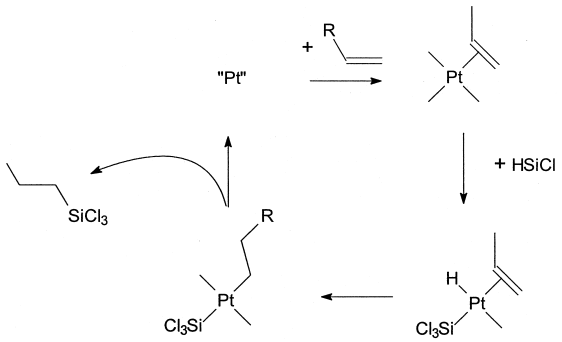


FIGURE 78 The Chalk-Harrod mechanism for hydrosilylation.

Instead of terephthalic esters the dimethyl esters are used and the condensation reactions are in part transesterifications. The monomers have to be extremely pure, because any monomeric impurity will end the growing chain; see Fig. 79.

The catalysts used for the transesterification reaction are simple salts, such as zinc or calcium acetate. For the direct esterification a mixture of Sb_2O_3 and titanium alkoxides are used. Similar salts are used for the manufacture of polyamides, polyurethanes, and polycarbonates. The mechanism involves interaction of the oxygen atoms of the monomers and the protons of the monomers with the anions added. Apparently a subtle balance between these interactions is needed in order to obtain effective catalysts. As for other reactions involving oxygenates or dioxygen, the knowledge about the various catalysts is empirical and a large variety of cations and anions have been tested. This is different from many reactions discussed above that involve discrete formation of metal-to-carbon bonds, via elementary steps for which we sometimes know how they can be influenced by ligand variation, such as the oxidative addition and reductive elimination steps.

SEE ALSO THE FOLLOWING ARTICLES

CATALYSIS, INDUSTRIAL • CATALYST CHARACTERIZATION • ELECTRON TRANSFER REACTIONS, GENERAL • KINETICS (CHEMISTRY) • ORGANOMETALLIC CHEMISTRY

BIBLIOGRAPHY

Beller, M., and Bolm, C., eds. (1998). "Transition Metals for Organic Synthesis; Building Blocks and Fine Chemicals," Wiley-VCH, Weinheim, Germany.

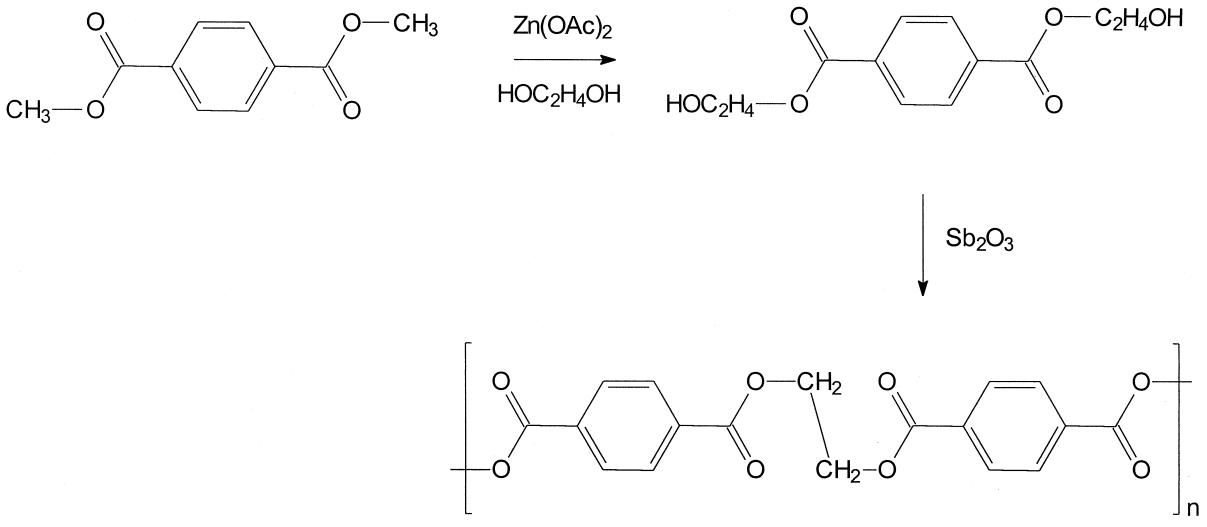


FIGURE 79 The two steps for polyester formation.

Cornils, B., and Herrmann, W. A., eds. (1996). "Applied Homogeneous Catalysis with Organometallic Compounds," VCH, Weinheim, Germany.
Cornils, B., and Herrmann, W. A., eds. (1998). "Aqueous Phase Organometallic Catalysis–Concepts and Applications," Wiley-VCH, Weinheim, Germany.
Diederich, F., and Stang, P. J., eds. (1998). "Metal-Catalyzed Cross-Coupling Reactions," Wiley-VCH, Weinheim, Germany.
Fink, R., Mülhaupt, G., and Brintzinger, H. H., eds. (1995). "Ziegler Catalysts," Springer, Berlin.
Noyori, R. (1995). "Asymmetric Catalysis in Organic Chemistry," Wiley, New York.
Parshall, G. W., and Ittel, S. D. (1992). "Homogeneous Catalysis," Wiley, New York.
van Leeuwen, P. W. N. M., and Claver, C., eds. (2000). "Rhodium Catalyzed Hydroformylation," In "Catalysis by Metal Complexes," Kluwer Academic Publishers, Dordrecht.
van Santen, R. A., van Leeuwen, P. W. N. M., Moulijn, J. A., and Averill, B. A., eds. (1999). "Catalysis, an Integrated Approach," 3rd ed., Elsevier, Amsterdam.

Fuel Chemistry

Sarma V. Pisupati
Pennsylvania State University

GLOSSARY

Atomization A process of breaking a liquid stream into fine droplets.
Combustion Rapid oxidation of fuels generating heat.
Equivalence ratio Ratio of the amount of air required for stoichiometric combustion to the actual amount of air supplied per unit mass of fuel.
Flame A chemical interaction between fuel and oxidant accompanied by the evolution of heat and light.
Fluidization A process of making a bed of particles behave like a fluid by a jet of air.
Fuel Any substance that can be used to produce heat and or light by burning.
Gasification Conversion of solid fuel into a combustible gas.
Liquefaction Conversion of solid fuel into liquid fuel.
Selective catalytic reduction A method used to selectively reduce nitrogen oxides to nitrogen using a catalyst.

ENERGY is the life-blood of a modern society. The quality of life of any society is shown to be directly proportional to the energy consumption of the society. In the past century (1900–2000), the population of the world and the United States has increased by 263 and 258%, respectively. The energy demand increased by a startling 16

TABLE I Total Primary Energy Consumption and Their Sources

1999 (Quadrillion Btus)		
Source	World	U.S.
Petroleum	152.20	37.71
Natural Gas	86.89	22.1
Coal	84.77	21.7
Nuclear	25.25	7.73
Hydro Electric	27.29	3
Renewable	2.83	4.37
Total	381.88	96.6

and 10 times, respectively. A vast majority of this energy (about 85%) comes from fossil fuels. These fuels—coal, oil, natural gas, oil shale, and tar sands were formed over millions years by compression of organic material (plant and animal sources) prevented from decay and buried in the ground. Table I shows the total primary energy use and the sources. Most of the fuels are used to generate heat and/or power. This chapter deals with the fuels origin, properties, and utilization methods and chemistry of these processes.

I. FUELS

Fossil fuels are hydrocarbons comprising of primarily carbon and hydrogen and are classified as solid, liquid, and gaseous based on their physical state. Solid fuels include not only naturally occurring fuels such as wood, peat, lignite, bituminous coal, and anthracites, but also certain waste products by human activities like petroleum coke and municipal solid waste. Approximately 95% of the coal mined in the United States is combusted in boilers and furnaces to produce heat and/or steam. The other 5% is used to produce coke for metallurgical uses. Liquid fuels are mostly produced in a refinery by refining naturally occurring crude oil, which include gasoline, diesel, kerosene, light distillates, and residual fuels oils. Each of these has different boiling range and is obtained from a distillation process. Gaseous fuels include natural gas, blast furnace gas, coke oven gas, refinery gases, liquefied natural gas, producer gas, water gas and coal gas produced from various gasification processes. Except for natural gas, most of the gaseous fuels are manufactured.

A. Origin of Fossil Fuels

Earth is a closed system with respect to carbon, and therefore carbon on this planet has to be used and reused. A total account of carbon in the world would explain fossil fuel formation. Global carbon cycle illustrates the fate carbon in the world. Carbon exists in the world in three major reservoirs: in the atmosphere as CO_2, in the rocks as CO_3^{--}, and in the oceans, which occupy two thirds of the planet's surface, as carbonate (CO_3^{--}) and bicarbonates (HCO_3^-). The CO_2 in the atmosphere has a vital role in the formation of fossil fuels. The CO_2 in the atmosphere reacts with water vapor in the presence of sunlight to form the organic matter and oxygen by *photosynthesis* reaction. The organic matter can be of microscopic plant (phytoplankton) or microscopic animal (zooplankton) or higher plants. The dead organic matter through *decay* reaction combines with oxygen and forms CO_2 and H_2O. This decay reaction is exactly the reverse of the photosynthesis reaction. Fossil fuels have formed by minimization or prevention of the decay reaction by possibly inundating the organic matter by water or covering by sediments. The organic matter, microscopic plants and animals, and higher plants, is chemically comprised of protiens, lipids, carbohydrates, glycosides, resins, and lignin. Of these, lignin is predominantly present in higher plants. Other components are predominantly present in zoo, phytoplankton, and algae (microscopic organic matter). Coal is a complex material composed of microscopically distinguishable, physically distinctive, and chemically different organic substances called macerals and inorganic substances called minerals.

During the transformation of organic matter to coal, there is a significant loss in oxygen and moderate loss of hydrogen with an increase in carbon content by initial aerobic reactions and subsequent anaerobic reactions. This process leads to the formation of kerogen. The organic matter, which is rich in algae, forms alginitic kerogen (type I kerogen); whereas the organic matter, which is rich in fatty acids and long-chain hydrocarbons, leads to the formation of liptinic or Type II kerogen. The organic matter consisting of lignin structure forms Type III kerogen and leads to the formation of coal. Formation of kerogen from the organic matter is by a process called "diagenesis," and the transformation of kerogen to fossil fuels is by a process called "catagenesis." The temperature in the earth's surface increases with depth at a typical rate of 10–30°C/km. As the temperature and pressure due to overburden increase, peat is transformed into lignite in about 30–50 million years. The primary reactions that are believed to occur during this transformation are dehydration, decarboxylation, and condensation leading to a loss of oxygen, hydrogen, and some carbon. Progressive transformation of lignite to subbituminous coal to bituminous coal and then to anthracite occurs in a time period ranging from 50 to 300 million years. The reactions responsible for these changes resulting in rapid loss of hydrogen are dealkylation, aromatization, and condensation. Formation of anthracites from bituminous coals

TABLE II Compositional Analysis of Various Solid Fuels

Fuel component	Wood[a]	Peat[a]	Lignite, Darco seam, TX[b]	Subbituminous coal[b]	Bituminous coal (hvAb), Upper Clarion seam[b]	Anthracite, Primrose seam[b]	Petroleum Coke[a]
Moisture (as rec'd)	48.00	—	32.60	27.12	1.73	3.77	5.58
Volatile matter (d.b.)[c]	72.80	75.00	67.39	47.56	39.41	3.71	10.41
Fixed carbon (d.b.)	24.2	23.1	21.34	38.12	51.68	82.38	88.89
Ash (d.b)	3.00	2.70	11.27	14.32	8.91	13.91	0.71
Heating value (d.b) (Btu/lb)	9,030	8,650	11,375	10,842	13,390	12,562	15,033
Carbon	55.00	8,650	74.90	73.88	83.67	96.65	88.64
Hydrogen	5.77	5.60	4.58	6.28	5.33	1.25	3.56
Nitrogen	0.10	0.70	1.75	1.22	1.46	0.78	1.61
Sulfur	0.10	0.17	0.78	1.78	4.82	0.52	5.89
Oxygen	39.10	40.10	18.78	18.62	9.54	1.32	0.30

[a] C, H, N, S, and O are given on a dry ash free basis.
[b] C, H, N, S, and O are given on a dry mineral matter free basis.
[c] Dry basis.

requires higher pressures and temperatures in excess of 200°C. The higher temperatures are encountered when there is a magma nearby in the ground, and high lateral pressures are encountered where land masses collide leading to the formation of mountains. Both of these scenarios lead to the formation of anthracites. The nature of the constituents in coal is related to the degree of coalification, the measurement of which is termed rank. Coals may be classified according to (1) rank, based on the degree of coalification; (2) type, based on megascopic and microscopic observations (physical appearance) that recognize differences in the proportion and distribution of various macerals and minerals; and (3) grade, based on value for a specific use. Typical properties of these solid fuels are shown in Table II.

II. PROPERTIES FOR UTILIZATION

Coals vary widely from place and to place, and sometimes even within a few feet in a particular seam because of the nature of the precursor material and the depositional environment. Therefore, coals and other solid fuels are analyzed for certain important properties for utilization (summarized in Table III).

Coal rank is usually determined from an empirical analysis called the proximate analysis and calorific value or optical reflectance of vitrinite. The proximate analysis consists of determination of moisture, volatile matter, and ash contents, and, by difference from 100%, the fixed carbon content of a coal. The American Society for Testing and Materials (ASTM) method of classification of coals uses the dry, mineral matter-free volatile matter to classify coals above the rank of medium volatile bituminous. For coals with greater than 69% volatile matter, the method uses the moist (containing natural bed moisture but not surface moisture), mineral matter-free calorific value to classify coals below the rank of high volatile bituminous. The moisture content is obtained by heating an air-dried coal sample at 105–110°C under specified conditions until a constant weight is obtained. The moisture content, in general, increases with decreasing rank and ranges from 1 to 40% for the various ranks of coal. The moisture content is an important factor in both the storage and the utilization behavior of coals. The presence of moisture adds unnecessary weight during transportation, reduces the available heat consuming latent heat of vaporization, and poses

TABLE III Important Properties for Utilization

Property	Factors affecting
Compositional analysis	Proximate analysis
	Ultimate analysis
Heating value	
Grindability	Coal rank
	Moisture
	Ash
Combustibility	Proximate analysis
	Surface area
	Porosity
	Petrographic Analysis
Inorganic constituents	Associated with the organic structure
	Discrete inorganic minerals

some handling problems. Volatile matter is the material driven off when coal is heated to 950°C in the absence of air under specified conditions, and is determined practically by measuring the loss of weight. It consists of a mixture of gases, low-boiling-point organic compounds that condense into oils upon cooling, and tars. Volatile matter increases with decreasing rank. In general, coals with high volatile matter are known to ignite easily and are highly reactive in combustion applications. The calorific value is the amount of chemical energy stored in the coal, which is released as thermal energy upon combustion and is directly related to the rank. The calorific value determines in part the value of a coal as a fuel for combustion applications. The ultimate analysis includes elemental analysis for carbon, hydrogen, nitrogen, and sulfur and oxygen on a dry ash free basis. Oxygen content in the fuel is obtained by subtracting the sum of the percentages of C, H, N, and S from 100.

The grindability of a coal is a measure of its resistance to crushing. The ball-mill and Hardgrove grindability tests are the two commonly used methods for assessing grindability. The Hardgrove method is often preferred to the ball-mill test because the former is faster to conduct. The test consists of grinding a specially prepared coal sample in a laboratory mill of standard design. The percentage by weight of the coal that passes through a 200-mesh sieve (screen with openings of three-thousandths of an inch or 74 μm) is used to calculate the Hardgrove Grindability Index (HGI). Two factors affecting the HGI are the moisture and ash contents of the coal. A correlation between the HGI and rank indicates that, in general, lignites and anthracite coals are more resistant to grinding (low indices) than are bituminous coals. The index is used as a guideline for sizing grinding equipment in the coal processing industry.

Inorganic constituents in coal upon oxidation produce ash. The inorganic constituents are present in the form of discrete mineral particles or bonded to the organic coal structure or sometimes water associated. The manner in which the inorganic constituents are present in raw coals determines the intermediate and final form of ash. Lower-rank coals (lignites and subbituminous) tend to have more organically associated inorganics compared to high-rank coals.

Combustibility is measured by proximate analysis and some empirical laboratory tests. Volatile matter is an indicator of ease of ignition and the fraction of coal that is burnt in gas phase. The higher the volatile matter the less is left as char that needs to be burnt. The combustion of char takes a longer time because of the heterogeneous reaction between carbon and oxygen. Combustion of high volatile coals therefore is easier to ignite and releases most of the heat closer to the burner. This requires an increase in the velocity of the fuel and air mixture jet to keep the flame at a distance from the burner tip. However, coals with lower volatile matter have a delay in ignition and therefore cause flame stability problems. These coals require additional design measures such as high swirl or a "bluff body" to promote recirculation of hot gases to stabilize the flame. Combustibility of coal is also characterized by bench-scale laboratory techniques such as drop tube reactor, entrained flow reactor, or thermogravimetric analyzer to determine the reactivity of fuels. One such method is used to determine a burning profile. This is widely used to determine the combustion behavior of an unknown fuel relative to a reference fuel. This method originally was developed by Babcock and Wilcox. A burning profile is a plot of the rate at which a solid fuel sample loses its weight as a function of temperature, when heated at constant rate in air. The burning profile is used to determine the temperatures at which the onset of devolatilization and char combustion, peak reation rate, and the completion of char oxidation occur. The temperatures range over which these take place also indicates the heat release rates and zones.

For a better understanding of behavior of coal in a furnace, pilot-scale combustion tests are performed to characterize ignition and flame stability, combustion efficiency, gaseous and particulate emissions, slagging and fouling, and the erosion and corrosion aspects of a fuel.

III. COMBUSTION OF COAL

Combustion is rapid oxidation of fuels generating heat. Combustion of fuels is a complex process the understanding of which involves chemistry (structural features of the fuels), thermodynamics (feasibility and energetics of the reactions), mass transfer (diffusion of fuel and oxidant molecules for a reaction to take place, and products away from the surface), reaction kinetics (rates of reactions), and the fluid dynamics (bulk movement of the fuel and combustion gases) of the process.

Combustion of solid fuels generally involves three steps—drying or evaporation of moisture, thermal decomposition and devolatilization, and oxidation of solid residue or char. Since solid fuels differ widely in terms of physical and chemical properties, a general qualitative discussion on the combustion behavior of a coal particle is given here. The duration and chemistry of each of these processes depend on the type of fuel burnt and the size of the particles, heating rate, furnace temperature, and particle density. For example, wood, peat, and lignite fuels contain a large percentage of moisture and the drying time is quite long. Also the volatile matter is quite high

and therefore the time required for heterogeneous combustion of solid residue (char) is shorter. The drying time of a small particle is the time required to heat the particle to the boiling point of the water and evaporate the water. This depends on the amount of water in the particle, particle size, and the heating rate of the particle. The rate of heat transfer to the particle depends on the temperature difference between the boiler furnace and the particle temperature.

Devolatilization of a fuel particle involves thermal decomposition of the organic structure, thereby releasing the resulting fragments and products. The rate of devolatilization depends on the rate of heat transfer to the particle to break weak bonds resulting in the primary decomposition of products. The primary products of decomposition travel from the interior of the particle to the surface through the pores of the solid. While doing so, the primary products, depending on their nature, could react with each other or with the char surface and result in secondary products or deposits on the walls of the pores. The devolatilization process begins at about 350°C and is a strong function of temperature. The products consist of water, hydrogen-rich gases and vapors (hydrocarbons), carbon oxides, tars, light oils, and ammonia. The product distribution for bituminous coal and lignite is shown in Fig. 1. These volatiles may be released as jets and play an important role in ignition and char oxidation steps. A carbon-rich solid product called char/coke also is produced. The composition of the products of decomposition depends on the type of the fuel, peak temperature, and rate of heating of the particle.

Combustion of char is a slower process compared to volatile combustion and is critical in determining the total time for combustion of a coal particle. The steps that are involved in the oxidation of char are

1. Diffusion of oxygen from the bulk to the char surface
2. Diffusion of oxygen from the surface to the interior of pores of the char
3. Chemisorption of reactant gas on the surface of the char
4. Reaction of oxygen and carbon
5. Desorption of CO and CO_2 from the char surface
6. Diffusion of CO and CO_2 to the surface through the pores
7. Diffusion of products from the exterior surface to the bulk of fluid

The properties of the char are important for its oxidation. Pulverized coal particles, during devolatilization, are known to swell by 5–15% depending of the coal type and heating rate. Various types of chars with varying physical structures are produced. The volatile matter when released

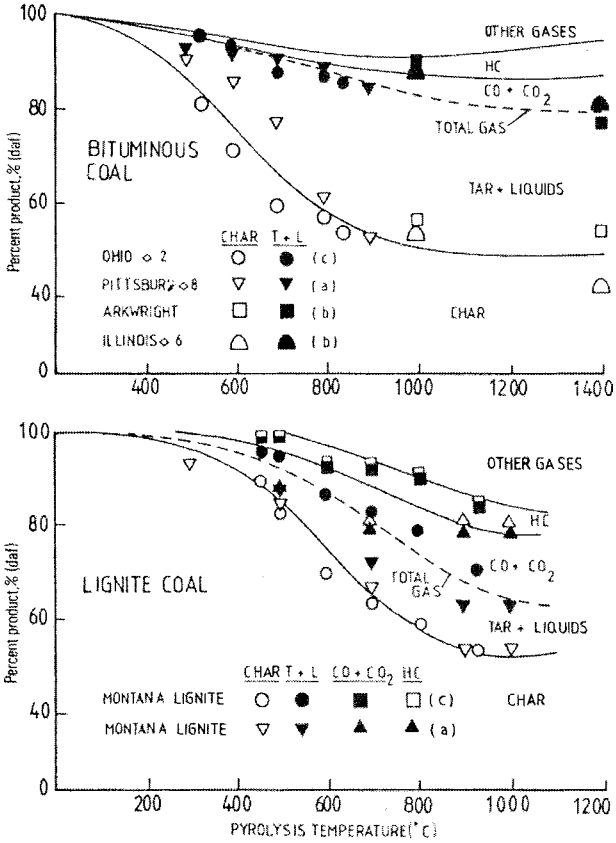


FIGURE 1 The product distribution for bituminous coal and lignite. From Lawn, C. W. (1987). "Principles of Combustion," Academic Press.

from a coal particle produces void space or porosity. The porosity of the char gradually increases as the conversion progresses. Ultimately the char particle becomes so porous that the particle fragments, as shown in Fig. 2. The combustion char that is produced can take place as surface reaction a (constant density and shrinking diameter) or a volumetric reaction (constant diameter and changing density). The surface reaction generally occurs when the temperatures are high and chemical reaction is fast and diffusion is the rate-limiting step, whereas volumetric reaction dominates when the temperatures are low and the oxidant has enough time to diffuse into the interior of the particle. The concentration profile of the reactant is shown in Fig. 2.

Ash formation occurs predominantly by two mechanisms. Volatile inorganic species are vaporized during char combustion and subsequently condense when the temperature is low downstream. The ash particles formed by this mechanism tend to be submicron in size; whereas, mineral inclusions come into contact with one another, and since the temperature in the pulverized coal combustion

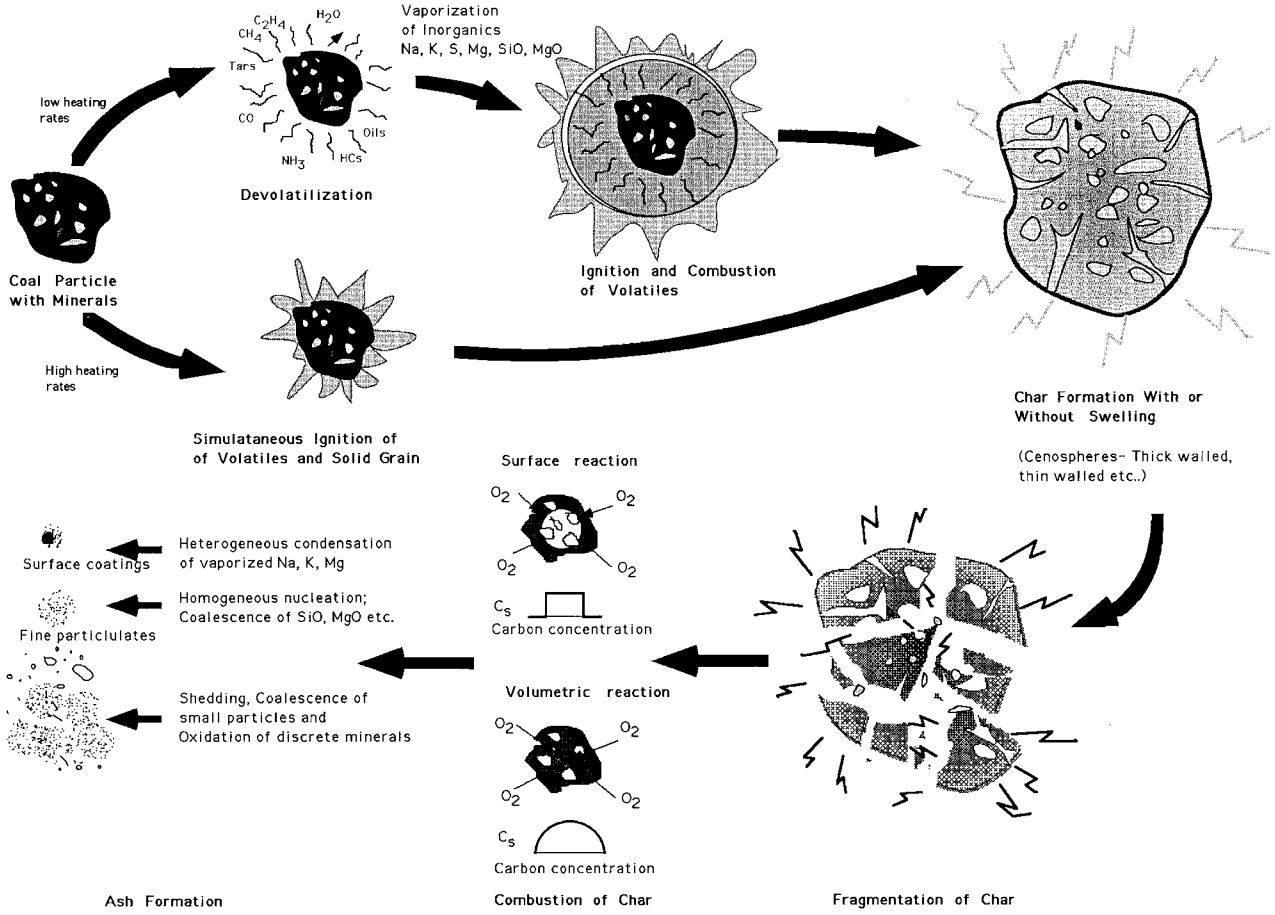


FIGURE 2 Simplified schematic of combustion of coal particles.

units is high enough to melt the molten ash, particles coalesce to form larger particles. Hence, the size distribution of pulerized coal ash tends to be bimodal.

A. Major Combustion Processes

Combustion methods of solid fuels are classified into four categories. These are fixed-bed combustion, fluidized-bed combustion, pulverized or suspension firing, and cyclone firing. These methods differ in the fuel size used and the dynamics of the fuel particles.

IV. FIXED BED COMBUSTION

Combustion of coal in fixed beds (e.g., in stokers) is the oldest method of coal use and it used to be the most common. Large-size coal pieces, usually size graded between 1/8 and 2 in., are supported on a grate and air is supplied from the bottom through the grate. The coal particles are stationary, hence the term "fixed" bed. Fixed-bed combustion systems can be further classified based on the coal feed system—overfeed, underfeed, spreader stoker, or traveling grate. Large size limits the rate of heating of the particles to about 1°C/sec and requires about 45 to 60 min for combustion. A schematic of a traveling grate combustion furnace is shown in Fig. 3. Coal particles initially devolatilize and the combustible volatile gases are burnt above the bed by the overfire air. The overfire air is crucial in obtaining complete combustion of the volatiles. In most of the stoker systems, coals which exhibit caking properties can be a problem due to clinker formation. Clinkers prevent proper air distribution combustion, the air leading to high unburnt carbon and higher carbon monoxide emissions.

When the fuel is drawn on to the grate from the coal hopper by the moving grate, the bed is approximately 10–15 cm. The coal particles initially require less air for drying and combustion. As the coal bed ignites and the plane of ignition travels down from the top to the grate, combustion air demand increases to a maximum. As the coal bed burns and reaches the other side, the demand for air decreases

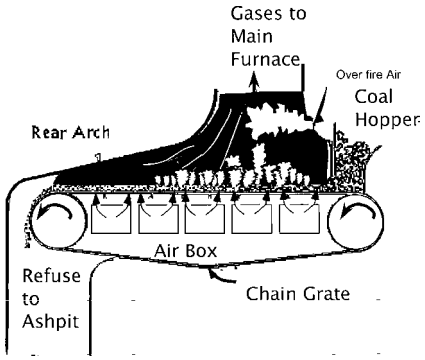


FIGURE 3 Schematic diagram of a traveling grate stroker.

to a minimum. Ash particles produced are dropped into a refuse pit on the other side of the combustion chamber. As shown in Fig. 4, the oxidant and the products of combustion have to travel through the layer of ash produced. This combustion method tends to produce a high amount of unburnt carbon and high(er) CO levels. Gaseous emissions from stoker systems are hard to control.

Since the coal bed height is constant, an increase in the throughput (larger size) can only be achieved by increasing the size of the grate. Due to this physical limitation of the grate size, stoker combustion units have an upper size limit of about 100 MMBtu/hr or 30 MW(thermal). Therefore, the application of this type of combustion method is limited to industrial and small on-site power generation units.

V. PULVERIZED COAL COMBUSTION

In the pulverized coal combustion system, coal ground to a very fine size (70–80% passing through a 200-mesh screen or 75 μm) is blown into a furnace. Approximately about 10% of the total air required for combustion is used to transport the coal. This primary air, preheated in the airheater, is used to drive out the moisture in the coal and transport the coal to the furnace. A majority of the combustion air is admitted into the furnace as secondary air, which is also preheated in the air preheater. The particles are heated at 10^5–10^6°C/sec, and it takes about 1–1.5 sec for complete combustion. This increases the throughput to furnace and, hence, it is the most preferred form of combustion method by electric utilities. Based on the burner configuration, pulverized coal combustion systems can be divided into wall-fired, tangentially fired, and down-fired systems. In wall-fired units the burners are mounted on a single wall or two opposite walls. Wall-fired systems with single wall burners are easy to design. However, due to uneven heat distribution, the flame stability and combustion efficiency are poor compared to the opposed wall-fired system. Because of the possibility of flame impingement on the opposite wall, this system may lead to ash slagging problems. An opposite wall-fired boiler avoids most of these problems by providing even heat distribution and better flame stability. A diagram of the opposed firing system is shown in Fig. 5. However, the heat release rate can be higher than single wall burner systems and could lead to slagging and fouling problems. A tangential firing system has burners installed in all four corners of the boiler (as shown in Fig. 6). The coal and primary air jets are issued at an angle, which is tangent to an imaginary circle at the center of the furnace. The four jets from four corners create a "fireball" at the center. A top view of the furnace is provided in Fig. 6. The swirling flame then travels up to the furnace outlet. This method provides the highest even heating flame stability. The temperature in the "fireball" can reach as high as 1700–1800°C. The burner is a key component in the design of the combustion system. It also determines the flow and mixing patterns, ignition of volatiles, flame stability, temperature, and generation of pollutants. Burners can be classified into two types—swirl and nonswirl. Primary air transports coal at velocities exceeding 15 meters/sec (the flame velocity) of the fuel. The flame velocity is a function of volatile matter. Secondary air ignites the fuel, determines the mixing pattern, and

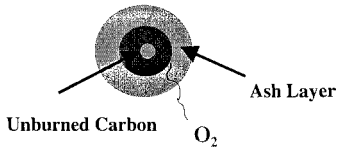


FIGURE 4 Combustion of a large coal particle.

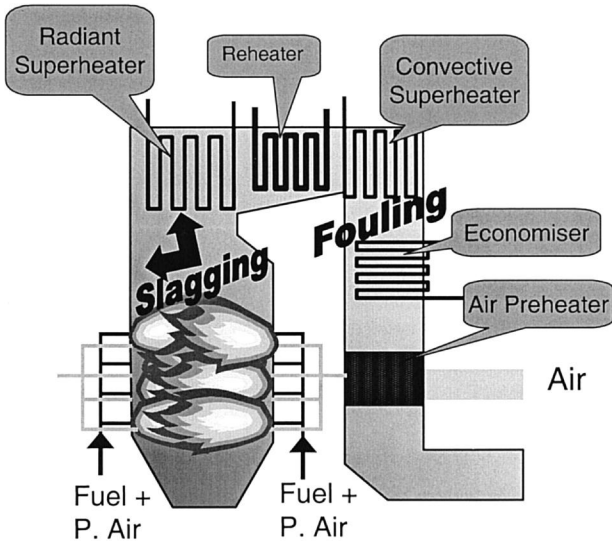


FIGURE 5 A schematic of a pulverized coal firing system.

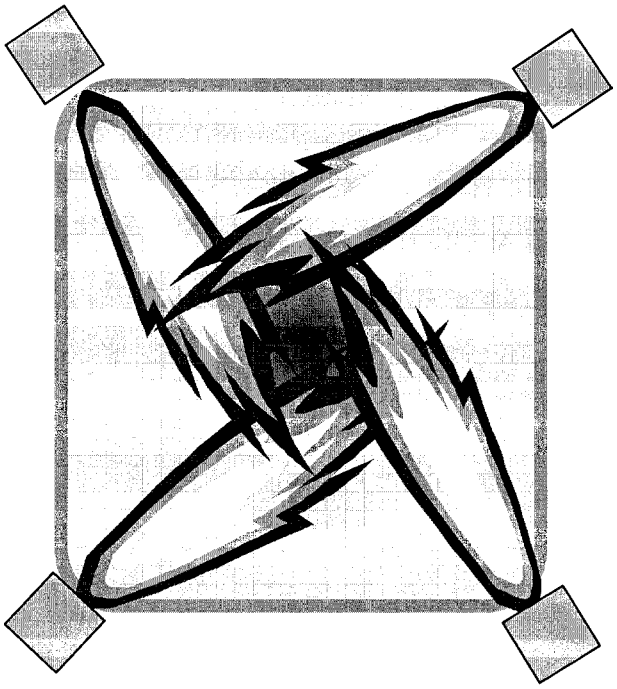


FIGURE 6 A plan view of a tangentially fired furnace.

also determines the NO_x formation. High volatile bituminous coals tend to have higher flame velocities compared to low volatile coals. Swirling flows promote mixing and esatablish recirculation zones. In swirling flows, the axial flux of angular momentum G and the axial flux of axial momentum Gx are conserved.

$$G_x = \int_{r_1}^{r_2} 2\pi r(p + \rho u^2)\, dr = \text{constant}$$

$$G_\phi = \int_{r_1}^{r_2} 2\pi \rho u w r^2\, dr = \text{constant},$$

where u and w are the axial and tangential components of the velocity at radius r, p is the static gage pressure, and r_1 and r_2 are radial limits of the burner. Swirl number is a measure of swirl intensity

$$S = \frac{G_\phi}{G_x r_b},$$

where r_b is the radius of the burner. A strong swirl usually has a value of $S > 0.6$. Stronger swirl is necessary when burning is difficult to ignite coals.

VI. FLUIDIZED BED COMBUSTION

Fluidization is a process of making solids behave like a "fluid." When a jet of air is passed though a bed of solids from underneath, the force due to drag pushes the particles up while the gravity pulls the particles down. When the velolity of the air increases above a certain point, the drag force exceeds the weight of the particles and the particles will start to fluidize. Under this condition, the solid bed material is made to exhibit properties such as static pressure per unit cross section, bed surface level and flow of solids through a drain etc., similar to that of a fluid in the bed. This fluidization behavior promotes rapid mixing of the bed material, thus promoting gas–solid and solid–solid heat transfer.

In a typical fluidized bed combustion process, solid, liquid, and or/gaseous fuels, together with noncombustible bed material such as sand, ash, and/or sorbent are fluidized. The primary advantage of this mode of combustion over pulverized and/or fixed-bed combustion is its operating temperature (800–900°C). At this temperature, sulfur dioxide formed by combustion of sulfur can be captured by naturally occurring calcium-based sorbents (limestone or dolostone). The reactions between the sorbent and SO_2 are thermodynamically and kinetically balanced in this range of temperatures. The operating temperature range of FBC boilers is low enough to minimize thermal NO_x production (which is highly temperature dependent) compared to other modes of combustion and yet high enough to achieve good combustion efficiency. In addition to these advantages, FBC boilers can be adapted to burn a wide range of fuels such as low-grade, low-calorific-value coal wastes, and high-ash and/or high-sulfur coals in an environmentally acceptable manner. Fluidized beds can be further subdivided into bubbling or circulating fluidized beds based on the operating characteristics, and into atmospheric and pressurized fluid beds based on the operating pressure.

Bubbling fluidized beds characteristically operate with a mean bed particle size between 1000 and 1200 μm (1.0 to 1.2 mm). The operating velocity in a BFBC boiler is above the minimum fluidizing velocity and less than a third of the terminal velocity (1–4 min/sec). Under the conditions of low operating velocity and large particle size, the bed operates in the bubbling mode, with a defined bed surface separating the densely loaded bed and the low solids freeboard region. In bubbling fluidized beds the goal is to prevent solids from carrying over, or elutriating, from the bed into the convective pass(es). Therefore, particles smaller than 500–1500 μm are not used in BFBCs so as to avoid excessive entrainment.

Circulating fluidized beds operate with a mean particle size between 50 and 1000 μm (0.05 to 1 mm). The hydrodynamics of the fast fluidized beds allow CFBC boilers to use sorbent particles as fine as 50 μm. CFBC boilers operate typically at about twice the terminal velocity of the mean size particles, ranging from 3 to 10 min/sec. This high operating gas velocity, recirculation rate, solids

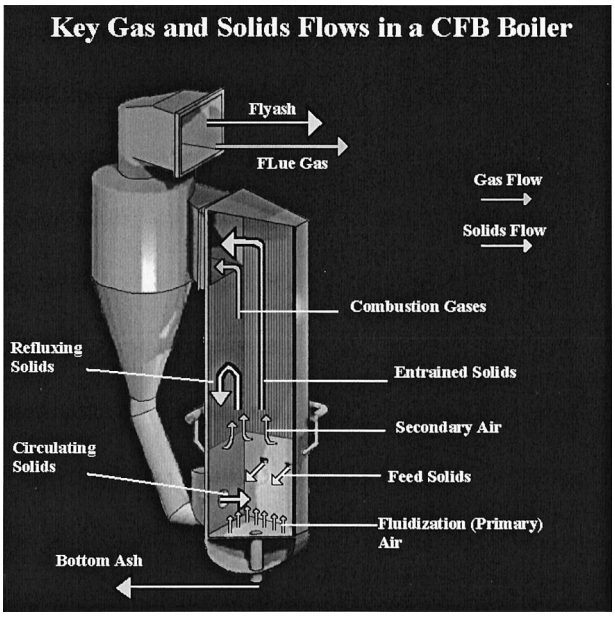


FIGURE 7 A schematic of a circulating fluidized bed boiler.

characteristics, volume of solids, and the geometry of the system create a hydrodynamic condition known as "fast bed" or "dilute phase refluxing." As a result of this condition solid agglomerates are continuously formed, dispersed, and reformed again. This motion also increases the slip velocity between the gas and solid phases, which increases the attrition or fragmentation of char and sorbent particles. The air required for the combustion is usually admitted in two streams: as primary and secondary air. In CFBC boilers, as shown in Fig. 7, the furnace can be conceptually divided into three sections: (1) the lower furnace zone, below the secondary air level where the solids flow regime is similar to a bubbling or turbulent bed and the particles range in size from approximately 1000 to 500 μm; (2) the upper furnace zone, above the secondary air level where fast fluidization occurs and the flow pattern approaches almost plug flow and particles from 500 to 150 μm are distributed in the circulating zone; and the (3) cyclone, where the material coarser than the cut point of the cyclone is recycled back into the combustor. The larger particles remain in the system for an extended period of time, ranging from hours for the dense phase in the lower zone to several minutes for the circulating phase in the upper zone of the furnace. Some of the fines make only one pass through the combustor and exit the cyclone, so that their residence time is similar to that of the gas.

The heating rate of a coal particle in an FBC boiler is approximately 10^3–10^4°C/sec depending on the particle size and the total average residence time of a particle in the combustion chamber is estimated to be 20 min.

Some of the advantages of CFB boilers include fuel flexibility (any fuel such as coal, peat, wood chips, agricultural waste, Refuse Derived Fuel (RDF), and oil, with sufficient heating value to rise the fuel and the combustion air to above the ignition temperature of the fuel can be successfully burnt in a CFB boiler), high combustion efficiency, smaller furnace cross section: good load following capabilities, higher turn down, and more environmentally friendly.

A. Behavior of Ash in Boiler Furnaces

The mineral matter in coal probably is the factor most responsible in considering coal a "dirty" fuel as compared to oil or gas. Because of the dominant role played by coal mineral matter in boiler performance, the need for a better understanding by R&D on the behavior of this impurity is now more important than ever. The most commonly observed problems associated with mineral matter in coal are slagging and fouling. Slagging is the most serious problem with pulverized-coal-fired steam generators, particularly wall slagging, the buildup of slag deposits on water wall tubes. This has several undesirable aspects:

1. It decreases heat transfer appreciably.
2. It can interfere with gas flow when massive deposits accumulate near burners.
3. It results in masses of slag that can cause physical damage to the furnace bottom when they become dislodged.
4. It plugs normal ash handling facilities with excessively large masses of bottom ash.
5. It provides a cover to encourage tube wastage by external corrosion.

The standard methods that are used to predict the slagging and fouling behavior depend on the ash analysis, which is reported as oxides of various inorganic elements. Slagging characteristics of coal ash can be predicted to some extent using the silica percentage. Generally, ash with a low silica percentage will cause the most problems with slagging, and coals with a high silica percentage will be the least troublesome. Thus coal benefeciation to control silica percentage provides one way of modifying slagging behavior, although the effectiveness of coal washing will depend on the chemical composition of the ash in the product. However, there are still a lot of factors that cause slagging (boiler operating parameters) that are not well understood.

Fouling of superheaters and reheaters causes problems such as plugging of gas passages, interference with heat transfer, and erosion of metal tubes. Of the factors leading to such deposits, the physical ones involving particle

motion, molecular diffusion, and inertial impaction are, to some extent, controllable by the design of the unit. The chemical factors associated with the ash, particularly the alkalies in the flue gas which are condensed on the surface of fly ash particles and the sintering characteristics of these particles depend upon the mineral species in the coal. Two alkalies, sodium and potassium, are generally held responsible for superheater and reheater fouling, with sodium given the most attention.

A fouling index has been developed to correlate sintering strength more closely with total ash composition:

$$\text{Fouling Index} = \text{Base/Acid} \times Na_2O,$$

where the base is computed as $Fe_2O_3 + CaO + MgO + Na_2O + K_2O$, and the acid as $SiO_2 + Al_2O_3 + TiO_2$ is expressed as weight percentage of the coal ash.

The fuel ash properties which boiler manufacturers generally consider important for designing and establishing the size of coal-fired furnaces include:

- The ash fusibility temperatures
- The ratio of basic to acidic ash constituents
- The iron/calcium ratio
- The fuel–ash content (mg/MJ)
- The ash firability

In addition to the furnace design, the ash content and characteristics also affect the superheater, reheater, and convective pass heat transfer surface design features. These characteristics along with a few others translate into the relative furnace sizes. It is recognized that the ash deposition is much better understood if the inorganic constituents in the coal are known rather than the ash composition after oxidation of the minerals. With the advancement in the computing power and image analysis in the late 1980s and early 1990s, Computer Controlled Scanning Electron Microscopy (CCSEM) has been a very useful tool in predicting ash behavior in furnaces. The determination of organically associated inorganic constituents is being made by chemical fractionation. These advanced characterization methods are now being more effectively used to help understand and solve the ash depositional problems in furnaces.

VII. ENVIRONMENTAL ISSUES IN COAL COMBUSTION

A. Sulfur Dioxide Emissions

Coal contains sulfur and nitrogen which vary typically between 0.5–5% and 0.5–2%, respectively. Sulfur upon combustion forms sulfur dioxide. About 65% of the total sulfur dioxide emissions are from electric utilities burning fossil fuels. Sulfur, 97–99%, in the fuel forms SO_2 and a small fraction of it is oxidized to SO_3. The equilibrium level of SO_3 in fuel lean combustion products is determined by the overall reaction

$$SO_2 + 1/2\,O_2 <=> SO_3.$$

The equilibrium constant ($Kp = 1.53 \times 10 - 5$ e 11,760/T atm $-1/2$) indicates that the SO_3 formation is not favorable at combustion temperatures. SO_3 concentration in the flue gas is typically 1–5% of the SO_2.

The SO_2 control techniques can be classified into precombustion, during combustion, and postcombustion methods.

1. Precombustion Control Methods

Precombustion control technologies can be fuel switching or fuel desulfurization. Since sulfur dioxide emissions are directly proportional to the amount of sulfur in the fuel, switching to a low sulfur fuel is a choice when permitted. However, fuel switching may not be an alternative if the regulations require a certain percentage of reduction in SO_2 emissions regardless of the fuel sulfur content. Fuel desulfurization involves reduction of sulfur from the fuels prior to combustion. Sulfur in coals is present in three forms: pyritic, organic, and sulfatic. Pyritic sulfur is the sulfur that is present as a mineral pyrite (FeS_2). The density of pyrite is higher than coal. Therefore, when coal is crushed and washed in water, the particles that are rich in the pyrite sink and the particles that are rich in carbon (organic) float. The sinks are rejected and the floats with less sulfur are used as clean coal. The degree of desulfurization depends on the distribution of sulfur in the coal. The finer the pyrite particles are, the more difficult the coal is to clean. These characteristics are of a coal that can be evaluated by washability curves based on a standard float and sink analysis. The organic sulfur is present in coal bonded to the organic structure of the coal. This form of sulfur is finely distributed throughout the coal matrix, and therefore is not possible to remove by physical coal cleaning.

2. During the Combustion Control Method

In an FBC system, limestones or dolostones are introduced into the combustion chamber along with the fuel. In the combustion chamber, limestones and dolostones undergo thermal decomposition, a process commonly known as calcination. The decomposition of calcium carbonate, the principal constituent of limestone, proceeds according to the following equation:

$$\begin{aligned} &CaCO_3\ CaO + CO_2 \\ &MgCa(CO_3)_2\ MgO + CaO + 2CO_2. \end{aligned} \quad (1)$$

Calcination, an endothermic reaction, occurs at temperatures above 760°C. Some degree of calcination is thought to be necessary before the limestone can react with gaseous sulfur dioxide. Calcined limestone is porous in nature due to the voidage (pores) created by the expulsion of carbon dioxide.

Capture of the gaseous sulfur dioxide is accomplished via the following reaction, which produces a solid product, calcium sulfate:

$$CaO + SO_2 + 1/2\,O_2\ CaSO_4. \quad (2)$$

The reaction of porous calcium oxide with sulfur dioxide produces a continuous variation in the physical structure of the reacting solid as the conversion proceeds. Because of the relatively high molar volume of $CaSO_4$ of CaO, the pore network within the reactant can be progressively blocked as conversion increases. For pure CaO prepared by the calcination of reagent grade $CaCO_3$, the theoretical maximum conversion of $CaCO_3$ to $CaSO_4$ has been calculated to be 57%. In practice, the actual conversion obtained using natural limestones is much lower due to the nature of the porosity formed upon calcination. Calcium utilizations as low as 15–20 mol% have been reported in some cases, although utilizations of about 30–40 mol% are typical. MgO will not react with sulfur dioxide at temperatures above 760°C; therefore, the sulfation reaction of dolomite is basically the reaction of sulfur dioxide with calcium oxide.

In pressurized fluidized bed combustion, however, the partial pressure of CO_2 is so high that calcination does not proceed because of thermodynamic restrictions. For example, at 850°C, calcium carbonate does not calcine if the CO_2 partial pressure exceeds 0.5 atmosphere. Under these conditions, the sulfation reaction is

$$CaCO_3\,(s) + SO_2 + 1/2\,O_2\ CaSO_4 + CO_2.$$

Increasing the pressure from one to five atmospheres significantly increases the sulfation rate and calcium utilization.

3. Flue Gas Desulfurization

Dry sorbent use in the case of pulverized coal combustion units is not efficient because of the operating temperature of the combustion chamber. At higher temperatures (>1100°C), CaO is known to sinter with a loss in the porosity and, therefore, the conversion. Also, at high temperatures, $CaSO_4$ is not stable and decomposes to CaO and SO_2 limiting this technology to FBC units.

Flue gas desulfurization systems in principle use alkaline reagents to neutralize the SO_2 and are classified as *throwaway* or *regenerative* types. This classification is based on the product fate. While, in a *throwaway* process the product produced by absorbing medium is thrown away (discarded), in a *regenerative* process, the SO_2 is regenerated from the product.

Most of the flue gas desulfurization systems operating in the United States use limestone or lime slurry scrubbing. In this system, limestone is finely ground (90% passing though a 325-mesh screen or 45 μm) and made into slurry. This slurry is finely sprayed in the absorption (scrubber) column. This slurry absorbs SO_2 in water as shown here

$$SO_2\,(g) + H_2O <=> SO_2 \cdot H_2O \text{ (dissolving gaseous } SO_2)$$

$$SO_2 \cdot H_2O <=> H^+ + HSO^{3^-} \quad \text{(hydrolysis of } SO_2).$$

This slurry with absorbed SO_2 is sent to a retention tank where the precipitation of $CaSO_3$, $CaSO_4$ and unreacted $CaCO_3$ occurs. Calcium carbonate has low solubility in water. Low pH promotes dissolution of $CaCO_3$ but low pH also lowers solubility of SO_2 in the scrubber. Therefore, a careful balance of pH is needed for this system. The following reactions take place in the retention tank, where

$$H^+ + CaCO_3 <=> Ca^{2+} + HCO^{3-1}$$
(dissolution of limestone)

$$Ca^{2+} + HSO^{3^-} + 2H_2O <=> CaSO_3 \cdot 2H_2O + H^+$$
(precipitation of calcium sulfate)

$$H^+ + HCO^{3-1} <=> CO_2 \cdot H_2O$$
(acid-base neutralization)

$$CO_2 \cdot H_2O <=> CO_2\,(g) + H_2O \text{ } (CO_2 \text{ stripping}),$$

the overall reaction being

$$CaCO_3 + SO_2 + 2H_2O\ CaSO_3 \cdot 2H_2O + CO_2.$$

B. Nitrogen Oxides

Nitrogen oxides, NO and NO_2, collectively known as NO_x, are formed during combustion in three ways. About 85–90% of the NO_x emitted from the combustion chamber is NO and 5 or 10% as NO_2. The three types of NO_x that form are by thermal, fuel, and prompt mechanisms.

Nitrogen oxide emissions from coal combustion can occur from three sources. Thermal NO_x primarily forms from the reaction of nitrogen and oxygen in the combustion air. The Fuel NO_x is a component that forms mainly from the conversion of nitrogen in the fuel to nitrogen oxides. Prompt NO_x is formed when hydrocarbon radical fragments in the flame zone react with nitrogen to form nitrogen atoms, which then form NO.

No definite rules exist to determine which nitrogen oxide formation mechanism dominates for a given stationary combustor configuration because of the complex interactions between burner aerodynamics and both fuel oxidation and nitrogen species chemistry. But in general, fuel nitrogen has been shown to dominate pulverized coal fired boilers, although thermal NO is also important in the postflame regions where over-fire air is used. Thermal NO contributions only become significant at temperatures above 2500°F in coal flames. Prompt NO formation is not typically an important mechanism during coal combustion. In the absence of fuel nitrogen, fuel NO is not a problem for natural gas flames. Prompt NO, however, is important in the vicinity of the inlet burners where reacting fuel fragments mix with the oxidizing air. Thus, in natural gas burners, both prompt and thermal NO contribute to the formation of nitrogen oxides.

C. Thermal NO

The principal reactions governing the formation of thermal NO are

$$N_2 + O \qquad NO + N$$

$$N + O_2 \qquad NO + O.$$

These two reactions are usually referred to as the thermal-NO formation mechanism or the Zeldovich mechanism. In fuel-rich environments, it has been suggested that at least one additional step should be included in this mechanism:

$$N + OH \qquad NO + H$$

The reactions (1), (2), and (3) are usually referred to as the extended Zeldovich mechanism. Experiments have shown that as the complexity of the reactors and fuels increase, it is difficult to evaluate the rate forms. The general conclusion is that although the nonequilibrium effects are important in describing the initial rate of thermal-NO formation, the accelerated rates are still sufficiently low that very little thermal NO is formed in the combustion zone and that the majority is formed in the postflame region where the residence time is longer.

D. Prompt NO

Prompt NO occurs by the fast reaction of hydrocarbons with molecular nitrogen in fuel-rich flames. This Fennimore mechanism accounts for rates of NO formation in the primary zone of the reactor, which are much greater than the expected rates of formation predicted by the thermal NO mechanism alone. NO measurements in both hydrocarbon and nonhydrocarbon flames have been interpreted to show that prompt NO is formed only when hydrocarbon radicals are available. It has been shown that the amount of NO formed in the fuel-rich systems is proportional to the concentration of N_2 and also to the number of carbons present in the gas phase. Hence, this mechanism is much more significant in fuel-rich hydrocarbon flames or in the reburning flames. Two reactions are believed to be the most significant to this mechanism:

$$CH + N_2 \qquad HCN + N \tag{3}$$

$$C + N_2 \qquad CN + N. \tag{4}$$

Reaction (4) was originally suggested by Fenimore in 1971. Reaction (5) is considered to be only a minor, but nonnegligible, contributor to prompt NO with its importance increasing with increasing temperature.

E. Fuel NO

Fuel NO is by far the most significant source of nitric oxide formed during the combustion of nitrogen-containing fossil fuels. Fuel NO accounts for 75 to 95% of the total NO_x accumulation in coal flames and greater than 50% in fuel oil combustors. The reason for fuel NO dominance in coal systems is because of the moderate temperatures (1500–2000 K) and the locally fuel-rich nature of most coal flames. Fuel NO is formed more readily than thermal NO because the N–H and N–C bonds common in fuel-bound nitrogen are weaker than the triple bond in molecular nitrogen which must be dissociated to produce thermal NO.

The main step, at typical combustion temperatures, consists of conversion of fuel nitrogen into HCN, step which according to some investigators (Fenimore, 1976; Rees *et al.*, 1981) is independent of the chemical nature of the initial fuel nitrogen. Once the fuel nitrogen has been converted to HCN, it rapidly decays to NH_i ($i = 1, 2, 3$) which reacts to form NO and N_2.

F. NO_x Control Technologies

In order to comply with the regulations for nitrogen oxides emissions, various abatement strategies have been developed. The most common methods used for NO_x control are air staging, fuel staging, flue gas recirculation, selective noncatalytic reduction (SNCR), or selective catalytic reduction (SCR). A determination of the most effective and least expensive abatement technique depends on specific boiler firing conditions and the emission standards. The principle of air staging is mainly reduction the level of available oxygen in zones where it is critical for NO_x formation. By doing so, the amount of fuel burnt at the peak temperature is also reduced. Air staging is

accomplished in the furnace by splitting the air stream for combustion. A part of the combustion air is introduced downstream as over-fire air (OFA), intermediate air (IA), or over-burner air (OBA). The principle of fuel staging or reburning involves reduction of the NO_x already formed in the flame zone by reducing it back to nitrogen during combustion. This is usually accomplished by injecting fuel into a second substoichiometric combustion zone in order to let the hydrocarbon radicals from the secondary fuel reduce the NO_x produced in the primary zone. To complete the combustion process of the reburn fuel, additional air is introduced downstream. However, the main drawback of this system is the short residence time available for the reburn fuel for complete oxidation. Therefore, this method uses mostly natural gas or any other highly reactive fuel so that combustion can be completed within the residence available.

In SNCR, chemicals are injected into the boiler, which then react with NO_x and reduce it to N_2.

Most commonly used chemicals are ammonia or urea. Other chemicals used in research work include amines, amides and amine salts, and cyanuric acid. Good mixing is essential for the success of this process and the optimum temperature window at which the reactions take place is 900–1100°C. This happens to be in a region where the heat transfer surfaces are present. Achieving good mixing is difficult.

When the limits of NO_x cannot be met by combustion control or by SNCR, SCR methods are used. NO_x concentration in the flue gas is reduced by injection of ammonia in the presence of a catalyst. The use of a catalyst reduces the optimum temperature window. The reaction products are water and nitrogen and this reaction is accomplished at much lower temperatures than SNCR (between 300 and 400°C). At lower temperatures, the unreacted ammonia can react with sulfur trioxide in the presence of water to form ammonium bisulfate (NH_4HSO_4), a sticky compound, which can cause corrosion, fouling, and blocking of downstream equipment.

This temperature is usually high enough to prevent condensation of (NH_4HSO_4). This system can reduce NO_x emissions by about 90–95%.

Most of the electric utilities are installing SCR systems to meet the current and future NO_x emission regulations.

VIII. COAL GASIFICATION

Gaseous fuels are easy to handle and use. They produce less by-products. As naturally occurring gaseous and liquid fuel resources deplete, synthetic gases and liquids produced from coal can act as suitable replacements. Coal gasification is a process of conversion of solid coal into combustible gases such as synthetic gas mixture (CO and H_2) or by methanation reaction conversion to synthetic natural gas (SNG). This is usually done by thermal decomposition, partial combustion with air or oxygen, and reaction with steam or hydrogen or carbon dioxide. Coal contains impurities such as sulfur, nitrogen, and minerals, which generate pollutants when burned. Most of these impurities and pollutant gases can be removed during the process of converting the coal into synthetic fuels. Therefore, gasification of coal is desirable from an environmental viewpoint and to overcome the transportation constraints associated with solid fuels. While the goal of combustion is to produce the maximum amount of heat possible by oxidizing all of the combustible material, the goal of gasification is to convert most of the combustible solids into combustible gases (such as CO, H_2, and CH_4) with the desired composition and heating value.

A. Gasification Reactions

During gasification, coal initially undergoes devolatilization (thermal decomposition) and the residual char undergoes some or all of the reactions listed here.

solid–gas

$$2C + O_2 = 2CO \text{ (partial combustion)}$$

$$C + O_2 = CO_2 \text{ (combustion)}$$

$$C + CO_2 = 2CO \text{ (Boudward reaction)}$$

$$C + H_2O = CO + H_2 \text{ (water gas)}$$

$$C + 2H_2 = CH_4 \text{ (Hydrogasification)}$$

gas–gas

$$CO + H_2O = CO_2 + H_2 \text{ (Shift)}$$

$$CO + 3H_2 = CH_4 + H_2O$$

For the thermodynamic and kinetic considerations, coal char is normally considered to be 100% carbon. The heats of reactions of the combustion or partial combustion reactions are exothermic (and fast), whereas most other gasification reactions (boudward, water gas, and hydrogasification) are endothermic and relatively slower. Usually, the heat requirement for the endothermic gasification reactions is met by partial combustion of some of the coal. It is also important to note that gasification reactions are sensitive to the temperature and pressure in the system. Equilibrium considerations for the reversible gasification reactions show that high temperature and low pressure are suitable for the formation of most of the gasification products, except for methane. Methane formation is favored at low temperature and high pressure. High temperatures tend to reduce methane, carbon dioxide, and water vapor

in the products. Coal char is considered to be a mixture of pure carbon and some inorganic impurities and structural defects. Certain impurities and structural defects are known to be catalytic; the absolute reaction rate depends on the amount and nature of the impurities and structural defects and also on physical characteristics such as surface area and pore structure. The physical structure controls the accessibility of gasification medium to the interior surface. These physical structural characteristics depend on the feed coal and the devolatilization conditions (heating rate and peak temperature).

There has been considerable interest in gasifying coals with several catalysts in order to minimize secondary reactions and produce the desired product gas distribution. However, the major problem with using catalysts is poisoning by various hydrocarbons and other impurities in coal. A major research effort in this area has been directed toward the gasification of carbon using alkali metals compounds. Potassium chloride and carbonate have been found to increase the amount of carbon gasified to CO and H_2. The same catalysts are found to reduce methane yield.

B. Classification of Gasification Processes

Gasification processes are primarily classified according to the operating temperature, pressure, reactant gas, and the mode of contact between the reactant gases and the coal/char. The operating temperature of a gasifier, usually dictates the nature of the ash removal system. Operating temperatures below 1000°C allow dry ash removal, whereas temperatures between 1000 and 1200°C cause the ash to partially melt, become "sticky," and form agglomerates. Temperatures above 1200°C result in melting of the ash and it is removed mostly in the form of liquid slag. Gasifiers may operate at either atmospheric or elevated pressure. Both temperature and pressure affect the composition of the final product gases. Gasification processes use one or a combination of three reactant gases: oxygen, steam, and hydrogen. The heat required for the endothermic reactions (heat absorbing) is supplied by combustion reactions between the coal and oxygen. Table IV illustrates the effect of gasification medium on the product species and the calorific value. Methods of contacting the solid feed and the gaseous reactants in a gasifier are of four main types: fixed bed, fluidized bed, entrained flow, and molten bath. The operating principle of fixed bed, fluidized bed, and entrained flow systems is similar to that discussed for combustion systems (see previous section). The molten bath approach is similar to the fluidized bed concept in that reactions take place in a molten medium (either slag or salt) with high thermal inertia and the medium both disperses the coal and acts as a heat sink, with high heat transfer rates, for distributing the heat of combustion. Table V summarizes some important gasification processes, conditions, and product gas compositions.

TABLE IV Effect of Gasification Medium on Products and Calorific Value

Gasification medium	Products	Calorific value of the products (MJ/m^3)
Air and steam	N_2, CO, H_2 and CO_2	5.6–11.2
Oxygen and steam	CO, H_2	11.2–14.5
Hydrogen	CH_4	35–38

C. Major Gasification Processes

The most important fixed-bed gasifier available commercially is the Lurgi Gasifier. It is a dry-bottom, fixed-bed system usually operated between 30 and 35 atmospheres pressure. Since it is a pressurized system, coarse-sized coal (25–37 mm) is fed into the gasifier through a lock hopper from the top. The steam–oxygen mixture (gasifying medium) is introduced through the grate located in the bottom of the gasifier. The coal charge and the gasifying medium move in opposite directions (counter-currently). The gasifier is operated at about 980°C and the oxygen reacts with coal to form carbon dioxide, thereby producing heat to sustain the endothermic steam–carbon and carbon dioxide–carbon reactions. The raw product gas consisting mainly of carbon monoxide, hydrogen, and methane leaves the gasifier for further cleanup. Besides participating in the gasification reactions, steam prevents high temperatures at the bottom of the gasifier so as not to sinter or melt the ash. Therefore, this gasification system is most suitable for highly reactive coals. Hot ash is periodically removed through a lock hopper at the bottom. Large commercial gasifiers measuring about 4 m in diameter and 6.3–8.0 m in height are capable of gasifying about 50 tons of coal per hour. Improved versions of the Lurgi Gasifier have been developed but not yet commercialized.

The Winkler gasifier is a fluidized-bed gasification system, which operates at atmospheric pressure. In this gasifier, crushed coal is fed using a screw feeder and is fluidized by the gasifying medium (steam–air or steam–oxygen mixture depending on the desired calorific value of the product gas) entering through a grate at the bottom. The coal charge and the gasification medium move cocurrently (in the same direction). In addition to the main gasification reactions taking place in the bed, some may also take place in the freeboard above the bed. The temperature of the bed is usually maintained at 980°C (1800°F) and the product gas consists primarily of carbon monoxide and hydrogen. The low operating temperature and pressure limit the throughput of the gasifier. Because of the low operating

TABLE V Summary of Some Gasification Processes

Process	Temperature (°C)	Pressure (Psia)	Gasification medium	Type of coal feed	Product gas composition[b] (dry basis)						Product gas calorific value[b] (Btu/scf)	Comments
					CO	CO_2	H_2	CH_4	C_2H_6	Others		
Lurgi[a]	980	400	O_2 + steam	Noncaking	68.5	29.5	40.4	9.4	—	1.2	302	
Winkler[a]	815	15	O_2 + steam	Any coal	33.4	20.5	41.9	3.1	—	1.1	275	
Synthane	980	1000	O_2 + steam	Any coal	16.7	28.9	27.8	24.5	0.8	1.3	405	
Hygas steam–oxygen	980	1000	O_2 + steam	Any coal	23.8	24.5	30.2	18.6	0.6	2.3	374	
Koppers–Totzek[a]	1510	20	O_2 + steam	Any coal	55.8	6.2	36.6	—	—	1.4	298	
			$_2$ + steam	Any coal	21.5	29.3	32.1	15.6		1.5	367	
Texaco	1480	600	O_2	Any coal in the form of coal–water slurry	46.6	11.5	38.7	0.7		2.7	300	
Cogas	980	75	Steam, and air as heat source	Any coal	7.4	9.3	26.2	34.0	8.3	14.8	726	
Producer	980	15	Air + steam	Any coal	29.0	4.0	12.0	3.0	—	52.0	130	
Water gas	980	15	Air + steam	Any coal	41	5.0	49.0	0.5	—	4.5	300	
Westinghouse	1930	20	Air + steam	Any coal	19.2	9.4	14.4	2.8		54.2	140	Oxygen can also be used
U-gas	1040	350	Air + steam	Any coal	9.8	12.0	10.3	—		67.9	75	
Combustion engineering	1930	20	Air	Any coal	24.4 60.6	4.1	10.7		—	—	125	Oxygen can also be used
CO_2 acceptor	815	150	Air + steam	Low rank	17.0 1.0	6.6	53.8		20.9	0.4	440	
Hydrane	815	>1000	H_2	Any coal	3.9	—	22.0	73.2	—	—	826	
Kellog salt	930	1200	O_2 + steam	Any coal	33.5	13.3	45.0	7.5	—	0.7	348	Oxygen can also be used

[a] Commercially available processes.
[b] Composition varies with coal used.

temperatures, lignites and subbituminous coals that have high ash fusion temperatures are ideal feedstocks for this type of gasifier. Units capable of gasifying 40–45 tons per hour are commercially available.

The Koppers–Totzek gasifier has been the most successful entrained-flow gasifier. This process uses pulverized coal (usually less than 74 μm) entrained (blown) into the gasifier by a mixture of steam and oxygen. The gasifier is operated at atmospheric pressure and high temperatures of about 1600–1900°C. The coal dust and gasification medium flow cocurrent (in the same direction) in the gasifier and because of the small coal particle size, the particle residence time is approximately 1 sec. Although this residence time is relatively short, high temperatures enhance the reaction rates, and therefore almost any coal can be gasified in this system. Tars and oils are evolved at moderate temperatures and crack at higher temperatures, and therefore, there is no condensible tarry material in the products and the ash melts and flows as slag in the K–T gasifier. Gasifiers with two diametrically opposite nozzles (also called heads) are most commonly used. However, the use of four nozzles doubles the throughput (over the two, nozzle design) and such configurations are also in use. The product gas is mainly synthesis gas (a mixture of CO and H_2) and is primarily used for ammonia manufacture. Since no heavy-duty moving parts are involved in this system, maintenance is minimal and availability is expected to be high.

D. Advanced Coal Gasification Systems

Many attempts have been made to improve the first-generation commercial gasifiers. British Gas Corporation has converted the Lurgi gasifier into a slagging type by increasing the operating temperature and thereby accommodating higher-rank coals which require higher temperatures for complete gasification. Another version of the Lurgi gasifier is the Ruhr-100 process with operating pressures about three times higher than that of the basic Lurgi process. Developmental work on the Winkler gasification process has lead to a pressurized version called the pressurized Winkler Process with an aim of increasing the yield of methane to produce Synthetic Natural Gas (SNG). Other processes in the developmental stages are the U-gas, Hy-Gas, Cogas, Westinghouse, and Synthane processes. The Texaco gasification system appears to be most promising entrained bed gasification system that has been developed. In this system coal is fed into the gasifier in the form of coal–water slurry and the water in the slurry serves as both a transport and a gasification medium. This system operates at 1500°C and the ash is removed as molten slag. Experience on demonstration units has indicated that the process has a potential to be used with combined-cycle plants for power generation. Another entrained bed gasification process under development is a pressurized version of the K–T process called the Shell–Koppers system.

These developing processes are primarily aimed at either increasing the operating pressure to increase the throughput and provide pressurized product gas for advanced power systems or increasing the operating temperatures to accommodate a variety of fuels, or both.

IX. COAL LIQUEFACTION

Similar to coal gasification, coal liquefaction is the process of converting solid coal into liquid fuels. The main difference between naturally occurring petroleum fuels and coal is the hydrogen to carbon molar ratio. Petroleum has approximately a hydrogen to carbon ratio of 2 where as coal has about 0.8. Coal also contains higher concentrations of oxygen, nitrogen, and significantly higher amounts of mineral matter than petroleum. Therefore, conversion of coal into liquid fuels involves hydrogenation (addition of hydrogen either directly or indirectly). Direct hydrogenation either from gaseous hydrogen or from a hydrogen donor solvent is termed direct liquefaction. If the hydrogen is added indirectly through an intermediate series of compounds, the process is called indirect liquefaction. In direct liquefaction processing, the macromolecular structure of the coal is broken down ensuring that the yield of the correct size of molecules is maximized and that the production of the very small molecules that constitute fuel gases is minimized. In contrast, indirect liquefaction methods break down the coal structure all the way to a synthesis gas mixture (CO and H_2), and these molecules are used to rebuild the desired liquid hydrocarbon molecules. This can be achieved by a variety of gasification techniques as discussed in the Coal Gasification section.

A. Liquefaction Reactions

Since coal in a complex substance it is often represented by an average composition and the reactions occurring during direct and indirect liquefaction can be illustrated as follows.

1. Direct Liquefaction

$$CH_{0.8}S_{0.2}O_{0.1}N_{0.01} + H_2 \longrightarrow (RCH_x) + CO_2, H_2S, NH_3, H_2O.$$

2. Indirect Liquefaction

$$CH_{0.8}S_{0.2}O_{0.1}N_{0.01} \longrightarrow CO + H_2 + H_2S, NH_3, H_2O \text{ (Gasification)}$$

$$CO + 2H_2 \longrightarrow CH_3OH \text{ (Methanol Synthesis)}$$

$$2CH_3OH \longrightarrow CH_3OCH_3 + H_2O \quad \text{(Methanol-to-Gasoline)}$$

$$nCO + 2nH_2 \longrightarrow (\text{–}CH_2\text{–})n + H_2O \quad \text{(Fischer–Tropsch Synthesis)}$$

$$nCO + (2n + 1)H_2 \longrightarrow (\text{–}CH_2\text{–})_{n+1} + H_2O.$$

Direct liquefaction of coal can be achieved with and without catalysts using pressures of 250 to 700 atmospheres and temperatures ranging between 425 and 480°C. In the indirect liquefaction process, coal is gasified to produce synthesis gas and cleaned to remove impurity gases and solids. The processes used to clean the gases depend on the impurities.

The principal variables that affect the yield and distribution of products in direct liquefaction are the solvent properties such as stability and hydrogen transfer capability, coal rank and maceral composition, reaction conditions, and the presence or absence of catalytic effects. Bituminous coals are the most suitable feed-stock for direct liquefaction as they produce the highest yields of desirable liquids, although most coals (except anthracites) can be converted into liquid products. Medium rank coals are the most reactive (react fast) under liquefaction conditions. Among the various petrographic components, the sum of the vitrinite and liptinite maceral contents correlates well with the total yield of liquid products.

3. Liquefaction Processes

The Bergius process was the first commercially available liquefaction process. It was developed during the First World War and involves dissolving coal in a recycled solvent oil and reacting with hydrogen under high pressures ranging from 200 to 700 atmospheres. An iron oxide catalyst is also employed. The temperatures in the reactor were in the range of 425–480°C. Light and heavy liquid fractions are separated from the ash to produce gasoline and recycle oil, respectively. In general, 1 ton of coal produces about 150–170 liters of gasoline, 190 liters of diesel fuel, and 130 liters of fuel oil. Separation of ash and heavy liquids, and erosion due to cyclic pressurization, posed difficulties which caused the process to be taken out of use after the war.

In the first generation, in a commercially operated, indirect liquefaction process called Fischer–Tropsch synthesis, coal is gasified first using the high-pressure Lurgi gasifier and the synthesis gas is reacted over an iron-based catalyst either in a fixed-bed Arge reactor or a fluidized-bed Synthol reactor. Depending on the reaction conditions, the products obtained consist of a wide range of hydrocarbons. Although this process was developed and used widely in Germany during the Second World War, due to poor economics it was discontinued after the war. This process is used in South Africa (Sasol) for reasons other than economics.

4. Developments in Liquefaction Processes

Lower operating temperatures are desirable in direct liquefaction processes since higher temperatures tend to promote cracking of molecules producing more gaseous and solid products at the expense of liquids. Similarly, lower pressures are desirable from an ease and cost-of-operation point of view. Recent research efforts in the area of direct liquefaction have concentrated on reducing the operating pressure, improving the separation process by using a hydrogen donor solvent (Consol Synthetic Fuels Process), operation without catalysts (Solvent Refined Coal (SRC)), and by using a solvent without catalysts (Exxon Donor Solvent) but using external catalytic rehydrogenation of the solvent. Catalytic effects in liquefaction are due to the inherent mineral matter in the coal and to added catalysts. Recent research efforts have focused on the area of multistage liquefaction to minimize hydrogen consumption and maximize overall process yields.

Later versions of Sasol plant (Sasol 2 and 3 units which also use indirect liquefaction process) have used only synthol reactors to increase the yield of gasoline and have reacted excess methane with steam to produce more CO and H_2.

Recent developments include producing liquid fuels from synthesis gas through an intermediate step of converting the synthesis gas into methanol at relatively low operating pressures (750–1500 psi) and temperatures (205–300°C). Methanol is then converted into a range of liquid hydrocarbons. The use of zeolite catalysts (as developed by Mobil) has enabled the direct production of gasoline from methanol with high efficiency.

X. LIQUID FUELS

Liquid fuels are obtained by refining naturally occurring crude oil. Like coal, crude oil from different places can differ in composition because of the precursor materials and the conditions for transformation organic matter to crude oil. Crude oil is a complex naturally occurring liquid containing mostly hydrocarbons and some compounds containing N, S, and O atoms. Crude oil consists of paraffins (straight-chain and branched-chain compounds), naphthenes (cyclo paraffins), and aromatics (benzene and its derivatives).

The average composition of crude oils from various parts of the world does not vary significantly. However, because of the variations in viscosity, density, sulfur, and

TABLE VI Typical Yield from a Barrel of Gasoline

Product	Yield (gallons)
Gasoline	19.5
Distillate fuel oil	9.2
Kerosene	4.1
Residual fuel oil	2.3
Lubricating oil, asphalt, wax	2
Chemicals for use in manufacturing (petrochemicals)	2

boiling points, these are separated into different fractions in a refinery. The most common refining operations are distillation, cracking, reforming alkylation, and coking. The demand for various products changes with the season and the lifestyle of the society. Typical yield from a barrel of crude oil is shown in Table VI.

XI. PROPERTIES FOR UTILIZATION

The majority of products (first four) are burnt in various devices. Gasoline is a mixture of light distillate hydrocarbons with a boiling range of 25–225°C consisting of paraffins, olefins, naphthenes, aromatics, oxygenates, lead, sulfur, and water. The exact composion varies with the season and geographic location. During summer months low volatile components are added to reduce the vapor pressure, whereas in winter months low boiling components are added to make it more volatile. Under the 1990 Clean Air Act Amendments (CAAA), the U.S. Environmental Protection Agency (EPA) developed reformulated gasoline (RFG) to significantly reduce vehicle emissions of ozone-forming and toxic air pollutants. RFG is required to be used in the nine major cities with the worst ozone air pollution problems. Similar to normal gasoline, RFG will contain oxygnates. Oxygenates increase the combustion efficieny and reduce emission of carbon monoxide. Table VII provides a comparison of properties for various gasolines.

Diesel fuel is also a mixture of light distillates but with higher boiling point components with a boiling point range of 185–345°C consisting of lower volatile and more viscous compounds. The average molecular weight is approximately 200.

Kerosene fuels are used in jet engines. Kerosene fuels have a wide range of boiling points. The aromatics in kerosene are limited due to their tendency to form soot.

Residual fuel oils are classified into five categoroies. Some of the important properties are listed here for good atomization.

1. Specific gravity—ratio between the weight of any volume of oil at 60°F to the weight of equal volume of water at 60°F
2. Viscosity—a measure of resistance to motion of a fluid
3. Flash point—The temperature at which the vapors generated "flash" when ignited by external ignition source
4. Pour point—The temperature at which the oil ceases to flow when cooled under prescribed conditions.

A. Combustion of Liquid Fuels

Fuel oil-fired furnaces, diesel engines, and distillate fuel-fired gas turbines utilize fine liquid sprays to increase the rate of evaporation and combustion rate of the fuel. In general the combustion of a liquid fuel takes place in a series of stages: atomization, vaporization, mixing of the vapor with air, ignition, and maintenance of combustion (flame stabilization). Recent advances have shown the atomization step to be one of the most important stages of liquid fuel combustion. The main purpose of atomization is to increase the surface area to volume ratio of the mixture. For example, breaking up of a 3-mm droplet into 30-μm drops results in 10^6 droplets. This increases the burning rate by 10,000 times. The finer the atomization spray the greater the subsequent benefits are in terms of mixing, evaporation, and ignition. The function of an atomizer is twofold: atomizing the oil and matching the momentum of the issuing jet with the aerodynamic flow in the furnace.

The atomizers for larger boiler burners are usually of the swirl pressure jet or internally mixed two fluid types, producing hollow conical sprays. Less common are the externally mixed two fluid types. The principal considerations in selecting an atomizer for a given application are turn-down performance and auxiliary costs.

There are differences in the structures of the sprays between atomizer types which may affect the rate of mixing of fuel droplets with the combustion air and hence the initial development of a flame.

For distillate fuels of moderate viscosity, (30 mm^2 sec^{-1}) at ordinary temperatures, a simple pressure atomization with some type of spray nozzle is most commonly used. Operating typically with a fuel pressure of 700–1000 kPa (7–10 atm) such a nozzle produces a distribution of droplet diameters from 10 to 150 μm. They range in design capacity of 0.5–10 or more, cm^3 sec^{-1}. A typical domestic oil burner nozzle uses about 0.8 cm^3 sec^{-1} of No. 2 fuel oil at the design pressure. Although pressure-atomizing nozzles are usually equipped with filters, the very small internal passages and orifices of the smallest tend to be easily plugged, even with clean fuels. With decreasing fuel pressure the atomization becomes

TABLE VII Properties of Various Types of Gasoline

	Fuel parameter values (national basis)				
	Conventional gasoline		Gasohol (2.7 wt% oxygen)	Oxyfuel	Phase I RFG
	Avg[a]	Range[b]	Avg	Avg	Avg
RVP3 (psi)	8.7-S 11.5-W[c]	6.9-15.1	9.7-S 11.5-W	8.7-S 11.5-W	7.2/8.1-S 11.5-W
T50 (°F)	207	141–251	202	205	202
T90 (°F)	332	286–369	316	318	316
Aromatics (vol%)	28.6	6.1–52.2	23.9	25.8	23.4
Olefins (vol%)	10.8	0.4–29.9	8.7	8.5	8.2
Benzene (vol%)	1.60	0.1–5.18	1.60	1.60 (1.3 max)	1.0
Sulfur (ppm)	338	10–1170	305	313 (500 max)	302
MTBE4[d] (vol%)	—	0.1–13.8	—	15 (7.8–15)	11
$EtOH_4$ (vol%)	—	0.1–10.4	10	7.7 (4.3–10)	5.7

[a] As defined in the Clean Air Act.
[b] 1990 MVMA survey.
[c] Winter (W) higher than Summer (S) to maintain vehicle performance.
[d] Oxygenate concentrations shown are for separate batches of fuel; combinations of both MTBE and ethanol in the same blend can never be above 15 vol% total.

progressively less satisfactory. Much higher pressures often are used, especially in engine applications, to produce a higher velocity of liquid relative to the surrounding air and accordingly smaller droplets and evaporation times. Other mechanical atomization techniques for production of more monodisperse sprays or smaller average droplet size (spinning disk, ultrasonic atomizers, etc.) are sometimes useful in burners for special purposes and may eventually have more general application, especially for small flows.

Conventional spray nozzles are relatively ineffective for atomizing of fuels of high viscosity such as No. 6 or residual oil (Bunker C) and other viscous dirty fuels. In order to transfer and pump No. 6 oil, it must usually be heated to about 373 K, at which its viscosity is typically 40 mm^2 sec^{-1}. Relatively large nozzle passages and orifices are necessary for the possible suspended solids. Dry steam may also be used in a similar way, as is common practice in the furnaces of power plant boilers using residual oil.

Combustion of fuel oil takes place through a series of steps, namely, vaporization, gasification, ignition, dissociation, and finally attaining the flame temperature. Vaporization or gasification of the fine spray of fuel droplets takes place as a physicochemical process in the combustion chamber. The temperature of vaporization for fuel oil is in the range of 100–500°F, depending on the grade of the fuel. Gasification takes place at about 800°F. The final flame temperature attained is between 2000 and 3000°F. The combustion of an oil droplet takes place in 2–20 msec depending on the size of the droplet. A typical characteristic of an oil flame is its bright luminous nature, which is due to incandescent carbon particles in the fuel-rich zone.

$$\tau = \frac{d_{p_p}^2}{\beta}.$$

Figure 8 illustrates the combustion of a single liquid droplet. Evaporation of liquid supplies the gaseous fuel

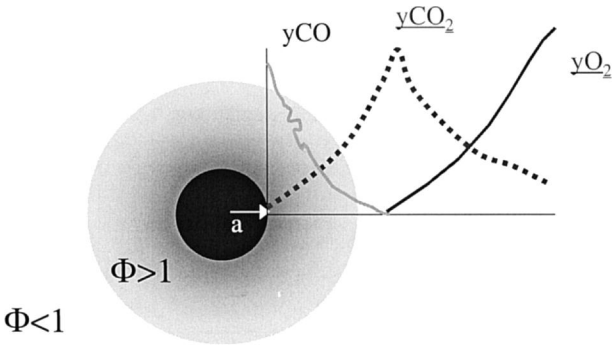


FIGURE 8 Evaporation and combustion of a liquid fuel droplet.

that burns in the gas phase. Evaporation is caused by heat transfer to the surface of the droplet. The time required for complete evaporation is given by

$$\beta = \frac{8\lambda}{\rho_l c_p} \ln(1 + B_T),$$

where, B_T is transfer coefficient, lambda is thermal Conductivity, $\rho_l =$ liquid density, and c_p is the heat capacity.

XII. MAJOR COMBUSTION METHODS

Internal combustion engines are devices that produce work using the hot combustion gases directly rather than steam. Three major types of IC engines are commonly used today using the top three products from a refinery. They are (1) spark engine (gasoline engine), (2) compression engine (diesel engine), and (3) gas turbine (aircraft engines).

A. Spark or Gasoline Engines

The most common engine is a 4-stroke engine. During the *intake* stroke, the fuel and air mixture is drawn into the cylinder with the exhaust valve closed. Then the air and fuel mixture is compressed in a *compression* stroke. At the top of the stroke, the spark plug ignites the mixture. During the *expansion* or power stroke, the high-pressure combustion gases expand moving the piston down and delivering the power. The gases expand completely, the exhaust valve opens, and the gases are expelled out during the *exhaust* stroke. The fuel and air are atomized and premixed in a carburetor. The higher the compression ratio the higher the efficiency of the engine. However, higher compression ratios also require higher octane number fuels. The octane number of a fuel is indicative of its antiknock properties. At equivalence ratios below 0.7 and above 1.4, the mixtures are generally not combustible. The equivalence ratio changes as the power requirement changes. For example, as the vehicle accelerates, high torque and power are required for which a fuel-rich mixture is used, whereas when the vehicle is cruising at high speeds the vehicle needs fuel lean mixtures. Therefore, in IC engines it is difficult to maintain the air to fuel ratio constant. Combustion in IC engines takes place in both oxygen-deficient and oxygen-rich environments, and the air and fuel mixtures are preheated by compression. Every time a fresh batch of fuel comes in flame is produced and quenched resulting in unsteady combustion. This results in continuous changes in the pollutant generation and emission.

B. Diesel Engines

Diesel engines are also IC engines. However, in Diesel engines, there is no carburetor. Only air is compressed to much higher pressures and the fuel is injected into the compressed air. As the fuel and air are mixed, the fuel evaporates and ignites (hence called compression ignition). The pressures used in the engines are almost twice those of the gasoline engines. Rate of injection and mixing of fuel and air determine the rate of combustion. Diesel engines are classified based on fuel injection, direct injection (DI), and indirect injection (IDI). Fuel quality is measured by cetane number (CN).

C. Gas Turbines

Another class of internal combustion engine is the gas turbine. Air is compressed to high pressures (10–30 atm) in a centrifugal compressor. Fuel is sprayed into the primary combustion zone where the fuel burns and increases the temerpature of the gases. The gas volume increases with combustion and the gases expand though a turbine. The power generated exceeds that required for the compressor. This drives the shaft to run an electric generator. In the aircraft applications, the gases are released at high velocity to provide the thrust. These systems are light weight compared to land-based systems. Land-based systems use either distillate oil or natural gas. Gas turbine-based power generation is used commonly to meet the peak power requirements rather than for base load operation.

D. Environmental Challenges for Liquid Fuel Utilization

Carbon monoxide is present in any combustion gas from any carbon containing fuel. The main factor that leads to its formation is incomplete combustion and in the IC engines continuous change from fuel-lean to fuel-rich conditions results in large emissions of CO. More than 70% of the CO emitted in the United States is from the transportation sector. CO emissions are also a function of vehicle speed. At lower speeds the emissions are higher. Cold starts also to contribute to higher emissions of CO. Oxygenates in the fuel aid complete combustion and result in a decrease of CO emissions. Catalytic converters placed at the end of the exhaust pipe oxidize the CO catalytically at lower temperatures.

Most of the hydrocarbon emissions are emitted through the exhaust. However, methane, ethane, acetylene, propylene, and aldehydes were found in the exhaust but were not present in the fuel. It can be deduced that these were formed during combustion. A significant amount of hydrocarbon emissions also come from the combustion chamber wall crevices and solid deposits. These hydrocarbon emissions reduce the NO_x emissions. However,

with the increase in the air to fuel ratio, CO and hydrocarbon emissions are reduced but NO_x emissions increase. To reduce all the emissions, three-way catalysts are being used. They not only oxidize CO and HCs but also reduce NO_x to N_2.

XIII. GASEOUS FUELS

A. Combustion of Gaseous Fuels

In any gas burner some mechanism or device (flame holder or pilot) must be provided to stabilize the flame against the flow of unburned mixture. This device should fix the position of the flame at the burner port. Although gas burners vary greatly in form and complexity, the distribution mechanism is fundamentally the same in most. By keeping the linear velocity of a small fraction of the mixture flow equal or less than the burning velocity, a steady flame is formed. From this pilot flame, the main flame spreads to consume the main flow at a much higher velocity. The area of the steady flame is related to the volume flow rate of the mixture:

$$\dot{V}_{mix} = A_f \times S_u$$

where,

$\dot{V}_{mix}$ = volumetric flow rate
A_f = area of the steady flame
S_u = burning velocity

The volume flow rate of the mixture is, in turn, proportional to the rate of heat input:

$$\dot{Q} = \dot{V}_{mix} \times HHV$$

where,

$\dot{V}_{mix}$ = volumetric flow rate
HHV = higher heating value of the fuel
$\dot{Q}$ = rate of heat input

In the simple Bunsen flame on a tube of circular cross section, the stabilization depends on the velocity variation in the flow emerging from the tube. For laminar flow (parabolic velocity profile) in a tube, the velocity at a radius, r, is given by:

$$v = \text{const}(R^2 - r^2)$$

where,

v = laminar flow velocity
R = tube radius
r = flame radius
const = experimental constant

Most of the commercial gas–air premixed burners are basically laminar-flow Bunsen burners and operate at atmospheric pressure—i.e., the primary air is induced from the atmosphere by the fuel flow with which it mixes in the burner passage leading to the burner ports where the mixture is ignited and the flame stabilized. The induced air flow is determined by the fuel flow-through momentum exchange and by the position of a shutter or throttle at the air inlet. Hence, the air flow is a function of fuel velocity as it issues from its orifice or nozzle, or of the fuel supply pressure at the orifice. With a fixed fuel flow, the equivalence ratio is adjusted by the shutter, and the resulting induced air flow also determines the total mixture flow, since desired air–fuel volume ratio is usually 7 or more, depending on the stoichiometry. Burners of this general type with many multiple ports are common for domestic furnaces, heaters, stoves, and for industrial use. The flame stabilizing ports in such burners are not always round and maybe slots of various shapes conform to the heating task.

Atmospheric industrial burners are made for a heat release capacity of up to 50 kJ/sec^{-1} and even despite their varied designs their principle of stabilization is basically the same as that for the Bunsen burner. In some the mixture is fed through a fairly thick-walled pipe or casting of appropriate shape for the application and the desired distribution of flame. The mixture issues from many small and closely spaced drilled holes, typically 1–2 mm in diameter, and burns, as rows of small pilot flame, spark or heated wire, usually located near the first holes, to avoid accumulation of the unburned mixture before ignition. The rate of total heat release for a given fuel–air mixture can be scaled with the size and number of holes—e.g., for 2-mm-diameter holes it would be 10–100 J/(hole) or in general 0.3–3 kJ cm^2 sec^{-1} of port area, depending on the fuel. The ports may also be narrow slots, sometimes packed with corrugated metal strips, to improve the flow distribution and lessen the tendency to flashback.

Gas burners that operate at high pressures are usually intended for much higher mixture velocity or heating intensity and the stabilization against blowoff must therefore be enhanced. This can be achieved by a number of methods such as (1) surrounding the main port with a number of pilot ports and (2) using a porous diaphragm screen.

In order to achieve high local heat flow the port velocity of the mixture should be increased considerably. In burners that achieve stabilization by causing pilot ports, most of the mixture can be burned at a port velocity as high as 100 Su to produce a long pencil-like flame, suitable for operations requiring a high heat flux.

SEE ALSO THE FOLLOWING ARTICLES

CARBON CYCLE • COMBUSTION • COAL PREPARATION • COAL STRUCTURE AND REACTIVITY • ENERGY FLOWS IN ECOLOGY AND IN THE ECONOMY • ENERGY RESOURCES AND RESERVES • FOSSIL FUEL POWER STATIONS—COAL

UTILIZATION ● FUELS ● INTEGRATED GASIFICATION COMBINED-CYCLE POWER PLANTS ● INTERNAL COMBUSTION ENGINES ● PETROLEUM REFINING ● POLLUTION, AIR

BIBLIOGRAPHY

Elliot, M. A. (ed.) (1980). "Chemistry of Coal Utilization," 2nd Supplementary Volume, Wiley, New York.

Borman, G. L., and Ragland, K. W. (1998). "Combustion Engineering," WCB/McGraw-Hill, Boston, MA.

Lawn, C. J. (ed.), (1987). "Principles of Combustion Engineering for Boilers" Academic Press, Harcourt Brace Jovanovich, New York.

Turns, S. R. (2000). "An Introduction to Combustion: Concepts and Applications," 2nd ed., Mc Graw-Hill Inc., New York.

Wen, C. Y., and Lee, E. S. (eds.) (1979). "Coal Conversion Technology," Addison-Wesley, MA.

Heinsohn, R. J., and Kabel, R. L. (1999). "Sources and Control of Air Pollution," Prentice-Hall, Upper Saddle River, NJ.

Williams, A., Pourkashanian, M., Jones, J. M., and Skorupska, N. (2000). "Combustion and Gasification of Coal," Applied Energy Technology Series, Taylor and Francis, New York.

Heterocyclic Chemistry

Charles M. Marson
University College London

GLOSSARY

Alkaloid Cyclic organic compound distributed in living organisms and containing nitrogen. Many alkaloids are bases of significant pharmacological activity.

Annelation Attachment, by synthesis, of a fragment to a structure of one or more rings to produce a bi- or polycyclic molecule. It can also refer to electronic changes that may occur when a structure is annelated.

Betaine Zwitterion of net neutral charge containing nonadjacent anionic and cationic sites. A hydrogen atom is not bonded to the cationic site.

Cyclodehydration Ring-forming reaction that proceeds with loss of water.

Degradation Procedure by which a molecule is cleaved to fragments, which are often identified, thereby assisting structural elucidation of the parent molecule.

Delocalization energy Calculated additional bonding energy that results when the assumption that electrons are constrained to isolated double bonds is removed. Corresponds to the resonance energy in valence bond theory.

Electrophile Electron-pair acceptor, such as H^+, $AlCl_3$, or Br^+.

Enantioselective A reaction that affords predominantly one enantiomer. Enantiomers of a compound are related to each other as nonsuperimposable mirror images.

Pi-excessive Heterocycle that contains a greater density of electronic charge than does a carbon atom of benzene. The converse holds for a pi-deficient heterocycle.

Mesoionic Five-membered heteroaromatic mesomeric betaine that cannot be satisfactorily represented by one polar structure and that possesses a sextet of electrons associated with the heterocyclic ring.

Nucleophile Electron-pair donor, such as RO^- or R_3N.

Protonation Formal addition of a proton (H^+).

Quaternization Alkylation of a tertiary amine to give a quaternary ammonium salt.

Resonance structures Two or more structures of the same compound that differ only in the pairing arrangement of the electrons. These structures have no differing chemical or physical properties; their identity is depicted by a double-headed arrow ($\leftrightarrow$) inserted between them. Resonance structures are nearly synonymous with canonical forms.

Tautomerism Structural isomers of different energies that are interconvertible by surmounting a low-energy

barrier. Isomerization involves the migration of a group or atom (especially H, referred to as proton tautomerism).

Ylid 1,2-Dipolar species, of net neutral charge, comprising a negatively charged carbon atom bonded to a positively charged heteroatom.

Zwitterion Species of net neutral charge that contains separate cationic and anionic sites.

HETEROCYCLIC COMPOUNDS possess a cyclic structure with at least two different kinds of atoms in the ring. By far the most common type, and that exclusively considered here, contains carbon together with one or more heteroatoms. The most common heteroatoms are nitrogen, oxygen, and sulfur, but many other elements can act as heteroatoms, such as phosphorus, tin, and silicon.

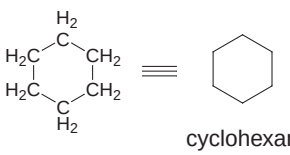
Heterocyclic compounds are extensively distributed in nature; many play crucial roles in the biochemistry of living organisms. For example, the genetic material deoxyribonucleic acid (DNA) contains a sequence of nitrogen heterocycles held together by hydrogen bonds across the heterocyclic rings. Many sugars exist in the form of five- or six-membered oxygen heterocycles. Most members of the vitamin B group are nitrogen heterocycles. Many naturally occurring pigments, antibiotics, and alkaloids are heterocyclic compounds. Owing to great advances in organic synthesis, many dyes, plastics, pharmaceuticals, pesticides, and herbicides contain heterocycles synthesized in the laboratory and not necessarily found in nature.

I. GENERAL ASPECTS OF HETEROCYCLIC SYSTEMS

A. Comparison with Carbocyclic Compounds

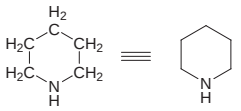
Molecules of organic chemical compounds are built up from a framework of carbon (C) atoms and associated hydrogen (H) atoms, to which oxygen (O) or other heteroatoms may or may not be attached. Because each carbon atom normally has four bonds to other atoms, carbon is said to be tetravalent. Hydrocarbons, which consist solely of carbon and hydrogen atoms, may be linear, branched, or cyclic. In the last case, such compounds are referred to as carbocyclic.

Cyclohexane (C_6H_{12}) is a typical carbocycle.



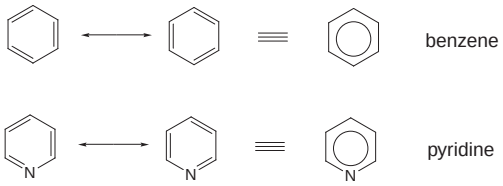
cyclohexane

A convenient method of depicting organic structures uses polygonal forms in which the carbon atoms reside at apices and are bonded to the appropriate number of hydrogen atoms if no other symbols (e.g., O for oxygen) are present. Replacement of a CH_2 group in cyclohexane by



piperidine

NH (nitrogen is usually trivalent) gives a heterocyclic compound called piperidine. It is often useful to compare a heterocycle to its carbocyclic analog; conversely, any number of heterocycles can be drawn by replacing the carbon atoms in carbocyclic compounds.



Removal of hydrogen (H_2) from a CH—CH bond results in the formation of a C=C bond, an unsaturated linkage referred to as olefinic. An aromatic structure contains alternating single and double bonds in a ring, such as benzene. Formal replacement of a CH in benzene by N gives the heterocycle known as pyridine. In general, the properties of aromatic and heteroaromatic compounds are quite different from those of their saturated (or partially saturated) counterparts. That is because a special stability is associated with those pairs of electrons (the so-called pi electrons) in a ring. Consequently, aromatic compounds are usually very stable and do not suffer addition of other atoms across the ring, unlike most other unsaturated compounds. A circle placed within a ring is used to denote a cyclic and delocalized array of *six* electrons in pi orbitals.

Saturated or partly unsaturated heterocycles usually resemble the acyclic (noncyclic) analogs closely in physical and chemical properties. Thus, piperidine has much in common with the acyclic amine diethylamine. Similarly, ethyl vinyl ether has much in common with 2,3-dihydrofuran.

diethylamine

2,3-dihydrofuran

ethyl vinyl ether

B. Heteroaromaticity

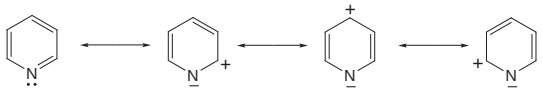
A nitrogen atom in a ring can be neutral or can carry a positive or a negative charge. Oxygen and sulfur atoms in a ring either are in the neutral form or carry a positive charge. For aromaticity to be observed in a heterocycle, electrons in an orbital perpendicular to the plane of the ring must overlap with the pi orbitals of other atoms in that ring. The lone pair of pyrrole is in such an orbital and the bonding can be expressed in the language of valence bond theory.

valence bond representation of pyrrole Z=NH

pi orbitals of pyrrole

Because substantial negative charge resides at all the ring carbon atoms of pyrrole (see valence bond structures, or canonical forms), it is referred to as a pi-excessive heterocycle. Owing to the excess negative charge at carbon, it is unsurprising that the chemistry of pi-excessive heterocycles is dominated by electrophilic attack.

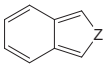
Pyridine is more aromatic than pyrrole, but for quite different reasons. Here, the three double bonds in the pyridine ring are markedly delocalized, and this accounts for the high degree of aromaticity of pyridine (see Table I).



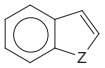
valence bond representation of pyridine

However, the lone pair on nitrogen is in an orbital parallel with the plane of the ring and hence cannot donate negative charge into it. Because nitrogen is a strongly electronegative (electron-attracting) element, the net result is to remove electronic charge from the ring, as just depicted in the canonical forms of pyridine. For this reason, pyridine undergoes electrophilic attack only under very forcing conditions and is referred to as a pi-deficient heterocycle.

Considerations similar to the above can be applied to all "aromatic" heterocycles, at least in principle. An interesting series is made up of the resonance energies of 3,4-benzoheterocycles, which are substantially lower than those of the 2,3-analogs, which retain much of the aromaticity associated with the benzene ring.



3,4-benzoheterocycle



2,3-benzoheterocycle

TABLE I Approximate Aromatic Stabilization Energies for Benzene and Some Heteroaromatic Rings (kcal/mole)

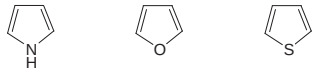
Benzene 36	Pyridine 31	Pyrazine 23
Pyrrole 20	Thiophene 25	Furan 18
Pyrazole 25	Imidazole 20	1,2,4-Triazole 20

The quantitative measurement and even the precise definition of aromaticity are problematic. One common measurement of aromaticity involves determining the differences in energy, by combustion, between a heterocycle and its carbocyclic analog; examples of values are given in Table I.

A second method involves measurement of magnetic ring currents induced in aromatic systems by external magnetic fields. A nuclear magnetic resonance spectrometer records the observed magnetic effect, which can be related to the degree of aromaticity.

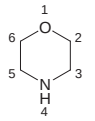
C. Nomenclature and Numbering of Heterocycles

Single heterocycles are extensively referred to by their well-known or "trivial" names, such as pyrrole, furan, and thiophene. Numbering of a monocyclic heterocycle, such as thiazole,

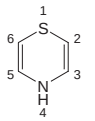


pyrrole furan thiophene

starts at the heteroatom of highest priority (oxygen taking priority over sulfur, which takes priority over nitrogen, etc.). Thus, morpholine and thiazine are numbered as shown.



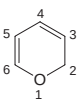
morpholine



1,4-thiazine

In all heterocyclic systems, the prefixes "oxa," "thia," and "aza" denote oxygen, sulfur, and nitrogen atoms, respectively. Also, two, three, or four identical heteroatoms bear the prefixes di, tri, and tetra, respectively, as in "dithia" and "triaza." Partially saturated rings are indicated by the prefix "dihydro" and "tetrahydro," etc., and/or 1*H*-, 2*H*-, etc., to identify saturated carbon(s) not taken into account by the prefixes of hydrogenation. Examples are tetrahydrofuran and 2*H*-pyran.

tetrahydrofuran



2*H*-pyran

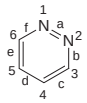
TABLE II Names of Monocyclic Heterocycles[a]

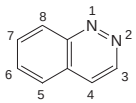
	One nitrogen in ring			One oxygen in ring		
Ring size	**Completely unsaturated**	**One double bond**	**Saturated**	**Completely unsaturated**	**One double bond**	**Saturated**
3	Az*irine*	Azirine	Az*iridine*	Ox*irene*	Ox*irene*	Ox*irane*
4	Az*ete*[b]	Azetine	Az*etidine*	Oxetium[b]	Oxetene	Ox*etane*
5	Pyrr*ole*	Pyrr*oline*	Pyrr*olidine*	Furan	Dihydrofuran	Tetrahydrofuran
6	Pyrid*ine*	Piperideine	Piperidine	Pyrylium[b]	Dihydropyran	Tetrahydropyran
7	Az*epine*	Dihydroazepine	Tetrahydroazepine	Ox*epine*	Dihydrooxepine	Ox*epane*

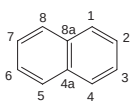
[a] Parts of names italicized are endings common to heterocycles of the given ring size.
[b] Cationic; other rings neutral.

Table II gives the common names of heterocyclic rings containing one nitrogen atom and, on the right-hand side, one oxygen atom in the ring. The complete name is that for the nitrogen or oxygen heterocycle, whereas that part underlined is the suffix for *any* ring containing one nitrogen atom and, on the right-hand side, any ring without nitrogen (oxygen has been chosen here). All the compounds listed contain trivalent nitrogen or divalent oxygen, except the oxetium and pyrylium rings, which contain cationic trivalent oxygen.

Systems containing two or more fused heterocyclic rings do not simply follow the rule of highest priority of each heteroatom for their numbering. Other things being equal, numbering is chosen so as to minimize the *sum* of the numbers assigned to the heteroatoms. Fusion of two or more heterocyclic rings is designated by lettering, starting with "*a*" between the 1,2 bond of the heterocyclic component. Thus, cinnoline *could* be referred to as benzo[*c*]pyridazine.


pyridazine

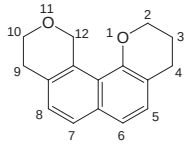

cinnoline


naphthalene

The numbering of cinnoline itself is based on that of the corresponding hydrocarbon, in this case naphthalene. An atom at a ring junction adopts the number of the atom that precedes it in a clockwise sense, such as 4, 4a, 4b. The use of heterocyclic components is the most common method of naming heterocycles. Many factors may need to be considered, in a designated system of priorities. A typical example that involves specifying unsaturation using "H" is 4*H*-1,3-dioxolo[4,5-*d*]imidazole.

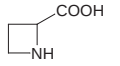

4*H*-1,3-dioxolo [4,5-*d*] imidazole

An important alternative to the "heterocyclic components" method is replacement nomenclature in which the *name* of a fused heterocycle is derived from that of the parent hydrocarbon. Thus, the following derivative of naphtho[1,2-*b*:8,7-*c*]dipyran can be referred to as an hydrogenated dioxabenzo[*c*]phenanthrene:


3,4,9,10-tetrahydro-2*H*,12*H*-1,11-dioxabenzo [*c*] phenanthrene

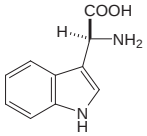
D. Occurrence and Isolation

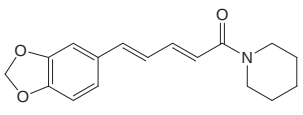
Both the number of sources in nature and the number of known, naturally occurring heterocycles are vast. Very simple heterocycles occur in nature, such a furan, found in certain low-boiling wood oils, and indole, present in jasmine and orange blossoms. The leaves of "lily of the valley" afford azetidine-2-carboxylic acid. Pyrrolidine occurs in carrot green.


azetidine-2-carboxylic acid

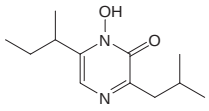

pyrrolidine

Tryptophan is an amino acid essential in human nutrition but not synthesized by the body; casein contains about 1% of tryptophan.

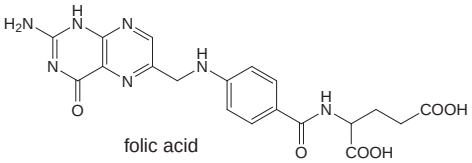

tryptophan


piperine

Piperine (*L*. piper = pepper) is a piperidine found in black pepper. Aspergillic acid is an antibiotic found in *Aspergillus flavus.* Folic acid is a B vitamin widely distributed in living matter.

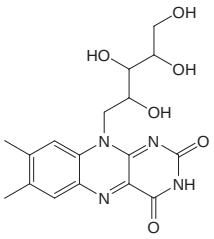


aspergillic acid



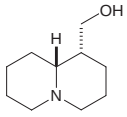
folic acid

More complex molecules frequently involve an array of chiral centers. Riboflavin (vitamin B_2) is an isoalloxazine involved in nutrition; its richest natural source is yeast.

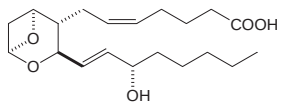


riboflavin

Other examples are 1-hydroxymethylquinolizidine, found in yellow lupin seeds, and thromboxane A_2, an important component in blood platelet aggregations.

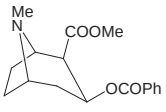


1-hydroxymethylquinolizidine



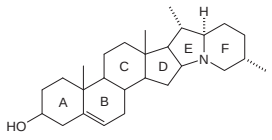
thromboxane A_2

Structural elucidation of many natural products paves the way for their synthesis; multistriatin was synthesized and used to trap the male elm bark beetle, which carries the fungus causing Dutch elm disease. Many naturally occurring heterocycles, such as cocaine (isolated from cocoa leaves) and morphine (from the opium poppy), have important pharmacophysiological properties.

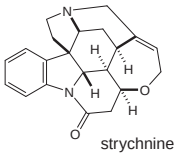


cocaine

Exceedingly complex structures have invariably proved fascinating to the organic chemist. Solanidine, found in potato shoots, is a formal fusion of a steroid skeleton (rings A–D) with a perhydroindole nucleus (rings E and F). Strychnine, the notorious poison, first isolated from *Strychnos ignatii* in 1818, contains several fused heterocyclic rings; it was the subject of a brilliant synthesis in the 1960s.



solanidine



strychnine

II. PHYSICAL PROPERTIES

Physical properties offer good criteria for the purity of heterocycles, just as they do for other organic compounds. The melting or boiling points were the characteristics formerly most frequently used, but optical spectra (based on absorption of light) and nuclear magnetic resonance spectra have become increasingly used. As for very many organic compounds, heterocycles display precise and readily reproducible physical properties.

A. Melting and Boiling Points

Melting and boiling points of common heterocycles are given in Tables III and IV. The boiling points of many of the heterocycles listed are substantially higher than those of the analogous cycloalkanes. Many simple heterocycles are mobile, pungent liquids at room temperature. Because molecular weight is related to vapor density, smaller molecules are, on the whole, more volatile than larger ones. An example can be seen in the comparison of alkylfurans with alkylbenzenes, the latter analogs being heavier by C_2H_2—O = 10 mass units; the furans boil at about 40°C *lower* than the corresponding benzenes.

TABLE III Boiling Points of Saturated Heterocycles and Corresponding Carbocycles (°C at 1 atm)

Ring size	Number and orientation of heteroatoms	Heteroatom type			Saturated cycloalkane
		NH	O	S	
3	One	56	11	55	−34
4	One	63	48	94	13
5	One	87	66	121	49
6	One	106	88	141	80
6	Two (1,2)	165[a]	—	—	80
6	Two (1,3)	150	106	—	80
6	Two (1,4)	145	101	200	80
7	One	138	120	174	119

[a] Corrected to atmospheric pressure by the method of Hass and Newton (1970). *In* "Handbook of Chemistry and Physics," 51st ed. (Weast, R. C., ed.), CRC, Cleveland, OH.

TABLE IV Melting and Boiling Points of Heteroaromatic Compounds Compared with Benzene (°C at 1 atm)

Ring system (with location of substituent)	Substituent										
	H	CH_3	C_2H_5	CO_2H	$CO_2C_2H_5$	$CONH_2$	NH_2	OH	OCH_3	Cl	Br
Benzene	80	111	136	122[a]	211	130[a]	184	43[a]	37[a]	131	155
Pyridine-2	115	128	148	137[a]	243	107[a]	57[a]	107[a]	252	171	193
Pyridine-3	115	144	163	235[a]	223	129[a]	65[a]	125[a]	179[a]	150	173
Pyridine-4	115	145	171	306[a]	219	156[a]	157[a]	148[a]	93[a]	147	174
Pyrrole-1	130	114	129	95[a,b]	180	166[a]	—[c]	—	—	—	—
Pyrrole-2	130	148	181	205[a,b]	39[a]	174[a]	—	—	—	—	—
Pyrrole-3	130	158	179	148[a]	78[a,b]	152[a]	—	—	—	—	—
Furan-2	31	64	92	133[a]	34[a]	142[a]	68[a]	80[a]	110	78	103
Furan-3	31	65	—	122[a]	179	168[a]	—	58[a]	—	80	103
Thiophene-2	84	113	133	129[a]	218	180[a]	214	217	154	128	150
Thiophene-3	84	115	135	138[a]	208	178[a]	—	—	—	136	157
Pyrazole-1	70[a]	127	137	—	212[a]	—	—	—	—	—	—
Pyrazole-3	70	205	209	214[a,b]	160[a]	—	285	164[a]	—	—	—
Isoxazole-3	95	118	138	149[a,b]	—	134[a]	—	—	—	—	—
Isoxazole-5	95	122	—	149[a]	174[a]	—	—	—	—	—	—
Imidazole-1	90	199	226	—	218[a]	—	—	—	—	—	—
Imidazole-2	90	141[a]	80	164[a,b]	—	—	—	250[a,b]	—	—	207[a]
Imidazole-4	90	56[a]	—	275[a,b]	157[a]	215[a]	—	—	—	—	130
Pyrimidine-2	123	138	—	270[a]	—	—	127[a]	320[a]	—	65[a]	—
Pyrimidine-4	123	141	—	240[a,b]	—	—	151[a]	164[a]	—	—	—
Pyrazine-2	57[a]	135	—	229[a,b]	—	189[a]	—	119[a]	187[a]	160	180

[a] Melts above 30°C.
[b] Melts with decomposition.
[c] A dash indicates that the compound is unstable unknown, or that the data are not readily available.

However, nitrogen heterocycles form a less regular series than do the furans because of significant hydrogen bonding and other intramolecular interactions.

Alkyl groups attached to heteroaromatic rings usually increase the boiling point in a regular manner: about 20–30°C for each methyl group and 50–60°C for each ethyl group. Where an NH group becomes substituted giving an NR group, a significant *decrease* in the boiling point occurs because of much diminished hydrogen bonding. The carboxylic acids and amides are all solids; a hydroxyl or an amino group in a nitrogen heterocycle usually results in a relatively high-melting solid.

B. Ultraviolet, Infrared, Nuclear Magnetic Resonance, and Mass Spectra

Ultraviolet (electronic absorption) spectra (Table V) give characteristic absorption bands for aromatic and many unsaturated heterocycles and can yield important information on their electronic structure. Absorption in the visible region (400–800 nm) occurs for extensive conjugated systems, especially if charge can be delocalized over several rings. This is the basis of many heterocyclic dyes.

Infrared spectra of heterocycles (and organic compounds in general) provide a molecular "fingerprint" of each compound (Table VI), in addition to allowing the recognition of many polar substituents, such as C═O and N─H groups.

Proton magnetic resonance spectra (Table VII) provide information on the number and nature of the environment of hydrogen atoms present in a molecule. Carbon magnetic resonance spectra (carbon isotope 13) allow each carbon atom in a molecule to be identified, together with the number of hydrogen atoms bonded to it. It is not surprising, therefore, that nuclear magnetic resonance has had an enormous impact on heterocyclic chemistry, allowing rapid structural identification, among many other important features. Mass spectrometers are capable of detecting many molecular ions, and ions formed by fragmentation, of heterocycles, even when only trace amounts are present.

C. Tautomerism

Tautomerism is very important in heterocyclic chemistry and many different types are known. Heteroolefinic compounds show tautomerism similar to their *acyclic* analogs.

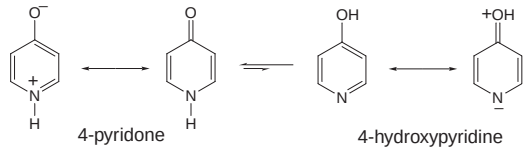
TABLE V Ultraviolet Spectral Maxima and Extinctions for Heteroaromatic Compounds

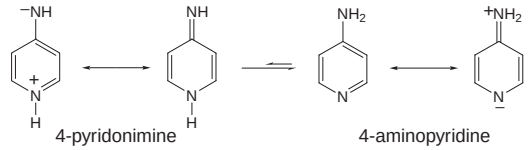
	Neutral form		Monoprotonated form			Neutral form		Monoprotonated form	
Compound	Wave length (nm)	Intensity ($\log_{10}\varepsilon$)	Wave length (nm)	Intensity ($\log_{10}\varepsilon$)	Compound	Wave length (nm)	Intensity ($\log_{10}\varepsilon$)	Wave length (nm)	Intensity ($\log_{10}\varepsilon$)
Pyrrole	210	4.20	241	3.90	Pyridine	257	3.42	256	3.70
Furan	208	3.90	—	—	Pyridazine	247	3.04	—	—
						300	2.51	238	3.21
Thiophene	231	3.87	—	—	Pyrimidine	243	3.51	242	3.64
Pyrazole	210	3.53	217	3.67	Pyrazine	261	3.77	—	—
						300	2.93	266	3.86
Isoxazole	211	3.60	—	—	Indole	270	3.77	280	3.68
					Quinoline	275	3.51	—	—
Isothiazole	244	3.72	—	—		299	3.46	313	3.79
						212	3.52	—	—
Thiazole	233	3.57	—	—	Isoquinoline	306	3.38	270	3.30
						319	3.47	332	3.63

The azoles show annular tautomerism of the type exemplified next.

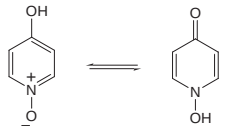
Hydroxyamino, mercapto, and methyl substituents are capable of tautomerism when attached to a heterocyclic ring containing a pyridine-like nitrogen atom. This type of tautomerism is found in the pyridines.

Pyridones predominate strongly over their tautomeric forms, the 2- and 4-hydroxypyridines. For the nitrogen analogs, the reverse is true, 2- and 4-aminopyridines being favored over the pyridonimine forms. This difference can be rationalized by considering the mesomerism of the alternative forms. The charge-separated form of 4-pyridone, being aromatic, is greatly preferred over the nonaromatic, charge-separated form of 4-hydroxypyridine. In the case of 4-aminopyridine, the charge-separated version of the imino form affords little stabilization, despite being aromatic, because nitrogen accommodates a negative charge much less readily than does oxygen.



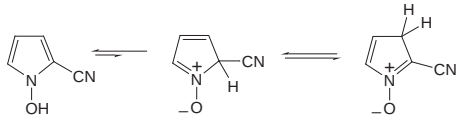


Hydroxypyridine 1-oxides are also tautomeric; 4-hydroxypyridine 1-oxide exists in approximately equal amounts of the pyridinium and 4-pyridone forms.



tautomerism of 4-hydroxypyridine 1-oxide

N-Hydroxypyrroles have, in principle, three tautomeric forms, but for 1-hydroxy-2-cyanopyrrole, only the *N*-hydroxy form has been observed.



tautomerism of 1-hydroxy-2-cyanopyrrole

For 2,3-dihydroxy-furan, -pyrrole, and -thiophene, the ketoenol forms are favored over the corresponding diketo forms.

tautomerism of 2,3-dihydroxyheterocycles (X = O, NH, S)

TABLE VI Some Characteristic Infrared Absorption Maxima for Ring-Stretching Modes of Common Heteroaromatic Compounds

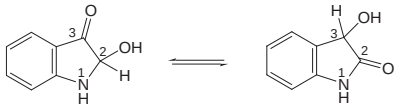
Compound	Absorption frequencies (cm^{-1})		
Pyrroles	1560–1530	1510–1480	1410–1390
Furans	1600–1560	1520–1470	1410–1370
Thiophenes	1540–1505	1440–1405	1370–1340
Pyridines	1610–1590	1580–1570	1485–1465

TABLE VII Chemical Shifts in Proton Resonance Spectra of Various Heterocycles

Compound	Chemical shift at given position[a]		
	2	3	4
Aziridine	1.48	—	—
Oxirane	2.54	—	—
Thiirane	2.27	—	—
Azetidine	3.58	2.32	—
Pyrrole	6.62	6.05	—
Furan	7.40	6.30	—
Thiophene	7.19	7.04	—
Pyridine	8.50	7.06	7.46
Pyridazine	—	9.17	7.68

[a] Parts per million on delta scale.

A somewhat similar system is 2,3-dihydroxyindole; here, the 2-hydroxy form is the sole tautomer at 20°C, but the equilibrium shifts progressively toward the 3-hydroxy form, which is favored by 95:5 at 100°C.



tautomerism of 2,3-dihydroxyindole

Whereas 2-aminothiophene exists in the amino form rather than the imino form, benzannelation can completely alter the tautomeric preference; thus, 1-aminobenzo[*c*]

tautomerism of 2-aminothiophene

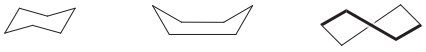
thiophene exists preferentially in the imine form, the amino form possessing appreciably less resonance energy.

tautomerism of 1-aminobenzo [*c*] thiophene

Many other tautomeric equilibria of heterocycles have been studied; theoretical calculations on the positions of these equilibria often agree with experimental values.

D. Conformation

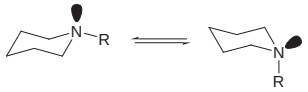
Whereas many fully unsaturated rings are very nearly planar, their saturated or partially saturated analogs containing tetrahedrally coordinated atoms are necessarily three-dimensional. This frequently gives rise to a preferred orientation of the atoms in a molecule, differing from other possible orientations (conformers) by rotation about single bonds. Thus, cyclohexane exists predominantly in the chair form, although boat and twist forms of cyclohexane are also known.



chair boat twist

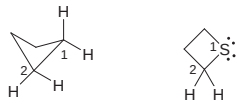
conformations of cyclohexane

With saturated heterocycles, conformational analysis is subject to additional considerations, such as the inversion of a lone pair on a heteroatom as for *N*-substituted piperidines, the effects of hydrogen bonding to a heteroatom, and dipole–dipole interactions involving a heteroatom (anomeric effects).



inversion at nitrogen of *N*-substituted piperidines

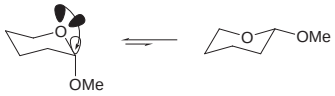
The geometries of the four-membered heterocyclic rings oxetane and thietane differ substantially from that of cyclobutane, the carbocyclic equivalent. Eclipsing



cyclobutane thietane

1,2-interactions are relieved by the puckered structure of cyclobutane. Similar eclipsing interactions are not significant between the sulfur and adjacent hydrogen atoms of thietane; consequently, this molecule is effectively planar.

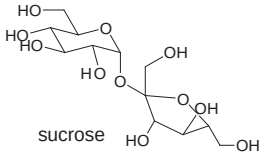
Like piperidine, tetrahydropyran adopts a chair conformation. Hydroxylated and alkoxylated tetrahydrofurans are important because such ring systems occur in various sugars and RNA. 2-Alkoxytetrahydropyrans, unlike alkoxycyclohexanes, exist preferentially with the alkoxy substituent in the axial position. In methoxylcyclohexane, for example, unfavorable 1,3-diaxial interactions make the equatorial position for the alkoxy substituent favored.



conformations of 2-methoxytetrahydropyran

However, for the corresponding tetrahydropyrans, the axial lone pair on the ring oxygen atom can donate electron density into the C─OMe antibonding (σ^*) orbital if the 2-alkoxy substituent is in the axial position. Accordingly, since such donation of electron density lowers the molecular energy, 2-alkoxy substituents exhibit a strong preference for the axial position (referred to as the

anomeric effect). The anomeric effect is a guiding principle in predicting the conformation of numerous carbohydrates.

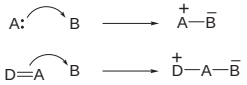


Sucrose ("sugar") is an important example of these stereoelectronic effects. There is little doubt that conformation is crucial to the function of many heterocycles found in nature.

III. CHEMICAL PROPERTIES

A. Chemical Reactivity

All chemical reactions involve the making and breaking of a bond or bonds. For each bond broken, one new one is made. A bond can be formed or broken in a reaction step that is either (1) ionic, (2) free radical, or (3) electrocyclic. Ionic mechanisms involve the transfer of two electrons (depicted by a curved arrow) from either a lone pair or a multiple bond to another atom:

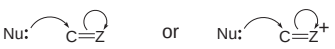


In a free-radical step, each atom contributes one electron (depicted by one single-headed arrow) to form a new bond:

In electrocyclic reactions, several bonds are formed and broken as part of a ring:

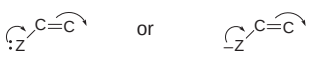
In heterocyclic chemistry, three patterns of reactivity are especially important:

1. The ability of a multiply bonded heteroatom to accept electrons at the site alpha to the heteroatom (Z):



The attack is easier when Z carries a positive charge.

2. The ability of a heteroatom attached to multiply bonded carbon to donate a pair of electrons:

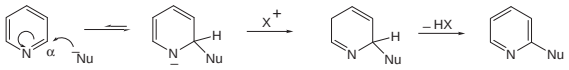


Again, this nucleophilic behavior is easier when Z carries a negative charge.

3. The tendency of heterocycles to revert to their initial degree of unsaturation if disturbed. Usually this means rearomatization.

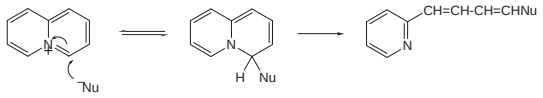
B. Nucleophilic Attack

Electron displacement toward the nitrogen atom often allows nucleophilic reagents to attack alpha to a nitrogen heterocycle, such as in the Chichibabin reaction, in which alkali-metal amides induce amination.



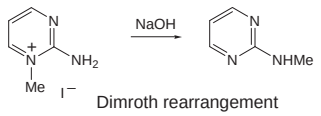
nucleophilic addition-elimination at Cα of pyridine

This type of reactivity is observed in benzenoid chemistry only when electron-withdrawing substituents are present. However, a powerful nucleophile is needed to form appreciable quantities of the anionic nitrogen species, which is of high energy, having been formed by dearomatization of the initial heterocycle. In the case of heterocyclic cations, nucleophilic attack is much easier, chiefly because of the resulting neutralization of opposite charges.



nucleophilic attack resulting in ring opening

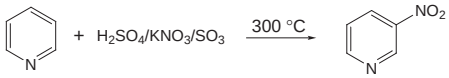
In many cases subsequent ring-closure occurs, to give a new heterocyclic ring. In certain cases, such as the Dimroth rearrangement of *N*-alkylated (or arylated) heterocycles to the corresponding alkylamine (or arylamine) heterocycles, the nature of the heterocyclic ring is not altered.



Dimroth rearrangement

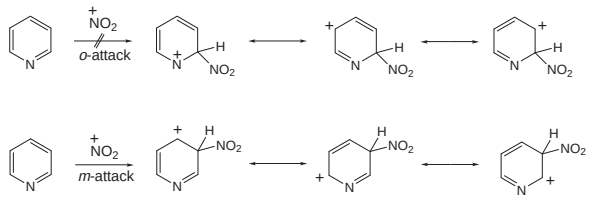
C. Electrophilic Attack

Electrophilic attack of pi-deficient heterocycles such as pyridine usually requires forcing conditions. Thus, nitration of pyridine at 300°C yields 3-nitropyridine in only poor yield.



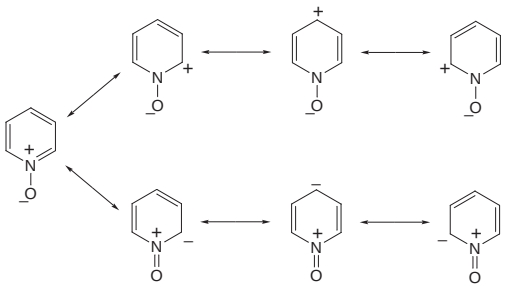
As with nucleophilic attack, the site of electrophilic attack can be rationalized, and is often predicted, by considering the relative stability of the structure formed prior

to the products. In terms of valence bond theory, the first drawn canonical form, formed by ortho attack on pyridine by nitronium ion, is of high energy because it carries a positive charge on the electronegative nitrogen atom. Hence this form contributes little to the overall stability of the cation formed by attack at the ortho position. The same arguments apply to attack at the para position. However, in the meta orientation all *three* forms contribute to lowering the energy of the cation; in practice the nitrated product is found to be almost exclusively *m*-nitropyridine. Much of the electrophilic and nucleophilic chemistry of heterocycles has been successfully rationalized along these lines.



attack of nitronium ion on pyridine

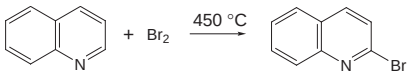
The canonical forms of pyridine *N*-oxide predict that the ring should be attacked by both nucleophiles and electrophiles. Both are indeed observed, such as the formation of 2- and 4-chloro- and 4-nitropyridine *N*-oxides, respectively.



canonical forms of pyridine *N*-oxide

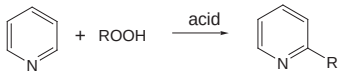
D. Radical Reactions

Halogenation of heterocycles at high temperature gives products in which the substitution pattern suggests that free radicals (rather than ionic halogen) are involved in their production.



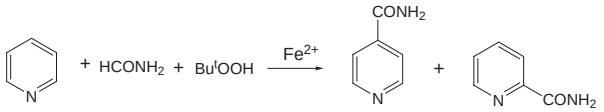
bromination of quinoline

Although neutral heteroaromatics often undergo alkylation by radical species in poor yield, if protonated heterocycles are used, good yields are frequently obtained. Under suitable conditions, pyridines, quinolines, and isoquinolines can be alkylated in the 2-, 2-, and 1-positions, respectively. Acylation of protonated pyridines,



alkylation of pyridine

quinolines, pyrazines, and quinoxalines via free radicals all proceed in fair yield. Thus, treatment of pyridine with free radicals generated from formamide gives isonicotinamide and the 2-isomer.

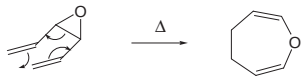


isonicotinamide

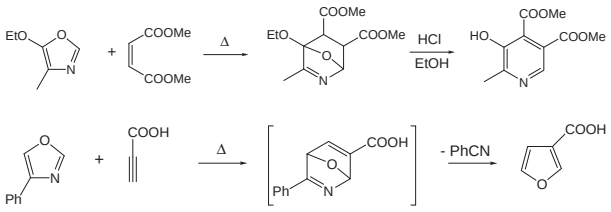
In marked contrast, homolytic arylation of heterocycles often gives low yields and is unselective. Phenyl radicals attack pyridine to give a mixture of 2-, 3-, and 4-phenylpyridines.

E. Electrocyclic Reactions

Electrocyclic reactions are very important both for the synthesis of heterocycles and in their transformation to new cyclic structures. Occasionally, intramolecular Diels–Alder reactions are used to form large heterocyclic rings.

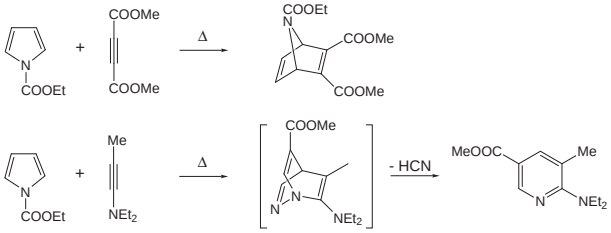


More often, two fragments undergo [4 + 2]cycloaddition to give a heterocyclic ring, as exemplified for the synthesis of pyridines and furans from oxazoles. Similar transformations include the conversion of pyrroles into azabicyclo



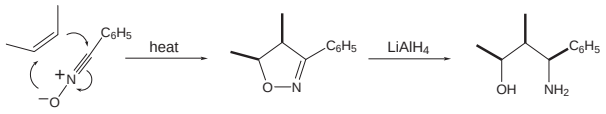
synthesis of pyridines and furans via cycloaddition reactions

compounds and pyridazines into pyridines.

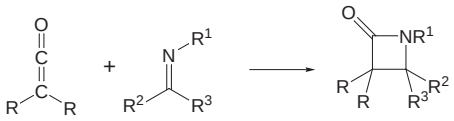


synthesis of azaheterocycles via cycloaddition reactions

1,3-Dipolar cycloaddition reactions are similar to Diels–Alder reactions, both proceeding through a cyclic transition state of six pi electrons, but differ in that three atoms supply the four-electron unit (in contrast to four atoms in a Diels–Alder reaction). Consequently, a five-membered ring results. Many cycloadditions, including 1,3-dipolar cycloaddition reactions, proceed with high stereocontrol. Sometimes the heterocyclic ring can be cleaved to give acyclic compounds of defined stereochemistry.

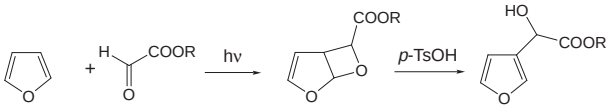


Four-membered heterocycles are often synthesized by [2 + 2]cycloadditions. An important example is the reaction of ketenes with imines to give beta-lactams.



[2 + 2] cycloaddition route to β-lactams

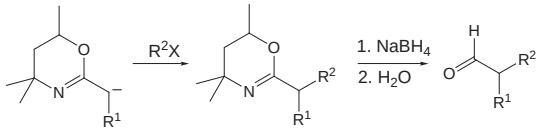
3-Substituted furans can be prepared by means of a photolytic cycloaddition. Many other [2 + 2]cycloadditions giving heterocycles are known.



synthesis of 3-substituted furans

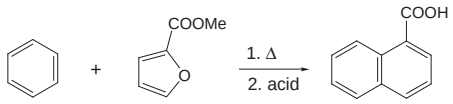
F. Use of Heterocycles in Synthesis

The following examples show how heterocycles can be used to prepare molecules that often cannot be readily synthesized by other routes. Thus, alkyl halides (R′X) are converted into alpha-substituted aldehydes via dihydro-1,3-oxazine intermediates.



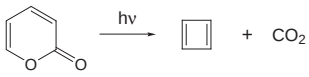
synthesis of aldehydes using dihydro-1,3-oxazines

Certain furans readily undergo Diels–Alder reactions to form adducts that can eliminate water to give substituted benzenes, and in certain cases benzene gives naphthalenes.



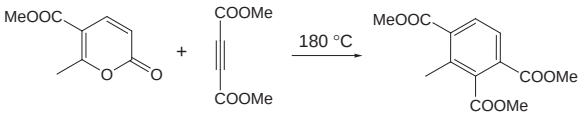
synthesis of naphthalene-1-carboxylic acid

Alpha-pyrone, on photolysis, yields the highly reactive cyclobutadiene.



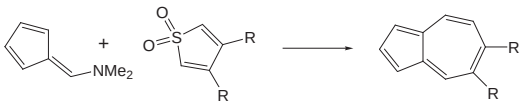
preparation of cyclobutadiene

Some alpha-pyrones can act as dienes, affording polysubstituted benzenes.



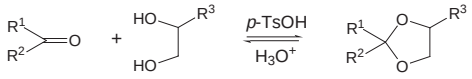
cycloaddition of an α-pyrone

Certain cyclic sulfones extrude sulfur dioxide when treated with *N*,*N*-dimethylaminofulvene, affording inky-blue azulenes.



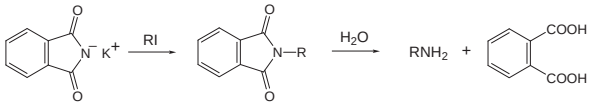
synthesis of an azulene

Heterocyclic rings can be used to protect functionality. Thus, the formation of cyclic ketals can be used to protect either ketones or aldehydes, *or* 1,2-diols.



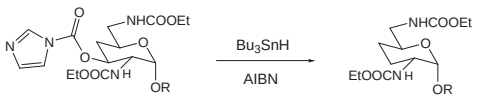
protection of functionality as a cyclic acetal

Deprotection occurs under mildly acidic aqueous conditions. The Gabriel phthalimide synthesis enables alkyl halides to be converted to primary amines. The action of ammonia itself on alkyl halides produces a mixture of amines and quaternary salts.

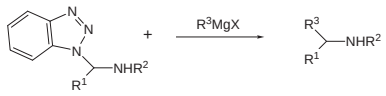


use of phthalimides in the Gabriel synthesis

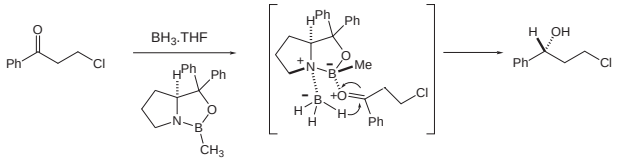
A general means of deoxygenating ROH into RH involves conversion of the hydroxy group into a thiocarbonylimidazolide followed by reductive cleavage using tributyltin hydride in the presence of the radical initiator azobisisobutyronitrile (AIBN).



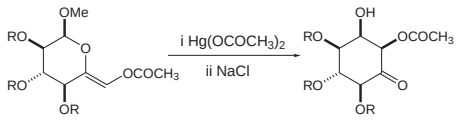
Benzotriazole ia an inexpensive and stable heterocycle that is soluble in a variety of common solvents. It is easily introduced into a molecule and can act as a satisfactory leaving group. Such properties enable its use in a wide variety of synthetic transformations, including displacements to form unsymmetrical secondary aliphatic amines.



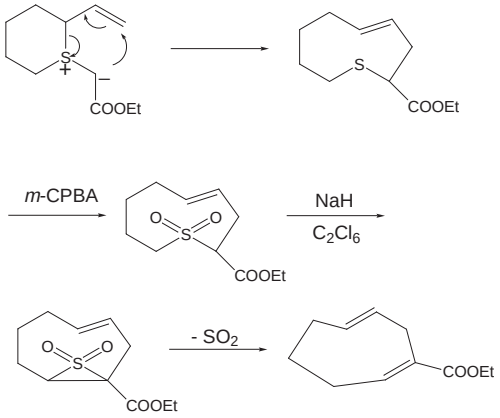
When borane becomes bonded to certain chiral oxazaborolidines, hydride transfer to ketones takes place, giving mainly one enantiomer of the alcohol. The larger group of the ketone points away from the heterocyclic ring, minimizing steric hindrance.



The ring opening of certain modified carbohydrates by mercury(II) acetate proceeds with subsequent cyclization (Ferrier rearrangement) to give a substituted cyclohexanone that can be reduced to an inositol derivative. Inositol phosphates are key messenger molecules in cellular signaling.

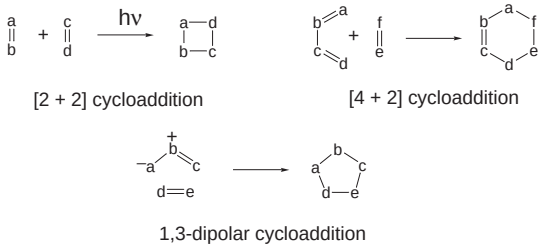


Sulfur heterocycles readily form sulfur ylides that can undergo electrocyclic ring expansion to give a medium-size ring. The sulfur can then be used to link two carbon atoms with extrusion of sulfur dioxide to form a cyclic alkene (Ramberg–Bäcklund reaction).

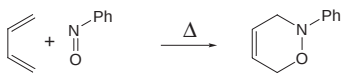


IV. SYNTHESIS OF HETEROCYCLES

The synthesis of heterocyclic rings has much in common with carbocyclic rings. Ring formation can be via cycloaddition, ring closure, ring expansion or contraction, ring-atom interchange (ANRORC reactions), or insertion of reactive intermediates into double bonds.

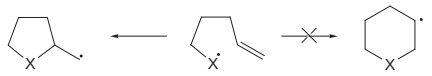


The Woodward–Hoffmann rules predict the regio- and stereochemistry of substituents in rings formed by cycloaddition reactions. In practice, a number of cycloaddition reactions are rarely or never observed, and these are forbidden by these important rules. The extremely useful Diels–Alder reaction is a [4 + 2]cycloaddition and was originally applied to carbocycles (atoms *a–f* = C), but it has been adapted to the synthesis of heterocycles.

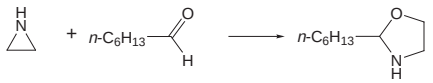


synthesis of a dihydro-1,2-oxazine

Ring-closure reactions follow well-established patterns, which are summarized in Baldwin's rules. For example, the ring closure of the terminal radical below leads to the so-called 5-*Exo*-Trig product rather than the 6-*Endo*-Trig product.



Ring expansion of aziridine to a tetrahydrooxazole is accelerated by the release of angle strain in the former.

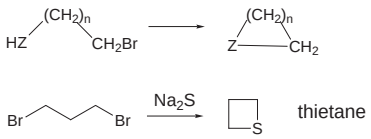


V. MAJOR CLASSES OF HETEROCYCLIC COMPOUNDS

A. Three- and Four-Membered Rings

1. Synthesis

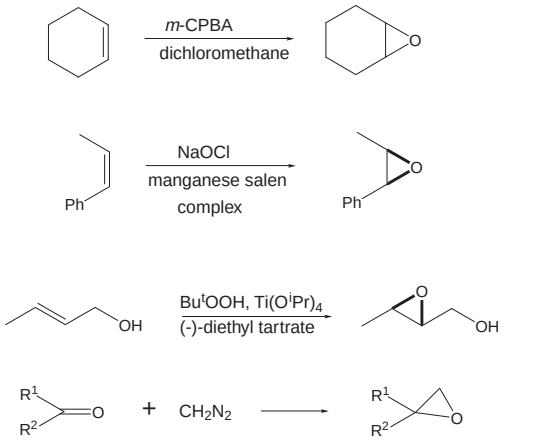
The saturated three-membered rings containing a nitrogen, oxygen, or sulfur atom are respectively referred to as aziridine, oxirane, and thiirane. They are usually readily formed by a ring closure that is entropically favored. The analogous formulation of four-membered rings ($n = 2$) is less favored and, unsurprisingly, less common.



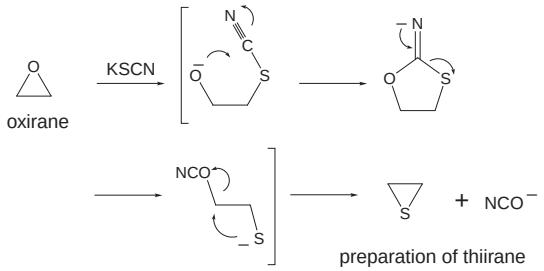
Epoxides (oxiranes) are extremely useful synthetic intermediates (see Section V.A.2 below). Most double bonds can be epoxidized by reaction with *m*-chloroperoxybenzoic acid (*m*-CPBA). Electron-deficient double bonds are an exception, but can be epoxidized using alkaline hydrogen peroxide.

Some alkenes undergo enantioselective epoxidation upon treatment with hypochlorite in the presence of a chiral manganese catalyst (Jacobsen reaction). The epoxidation of allylic alcohols by means of an alkylhydroperoxide attached to a titanium alkoxide–tartrate system is known as the Sharpless asymmetric epoxidation and is one of the most general and widely used asymmetric processes known; high enantioselectivity is usually observed.

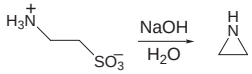
Carbonyl compounds react with diazoalkanes with loss of nitrogen to give epoxides. Indeed, addition of a carbene (R_2C:) or a nitrene (RN:) (or a carbene or nitrene equivalent) to a double bond is another general route to three-membered heterocycles.



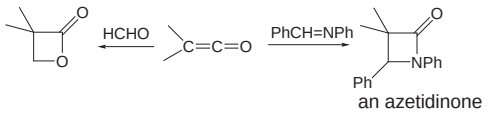
The conversion of oxirane into thiirane proceeds via as S_N ANRORC mechanism.



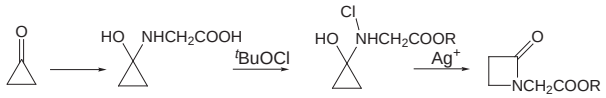
The formation of aziridine is illustrated by the following sequence.



Important routes to four-membered rings containing one (or more) heteroatoms are [2 + 2]cycloadditions, as previously discussed. Ring expansion of cyclopropanone

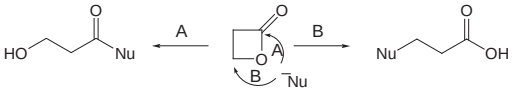


to azetidinones is also practical.



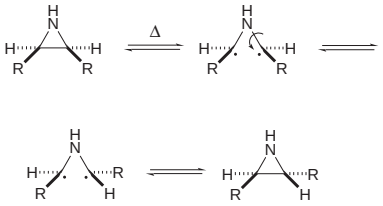
2. Properties and Importance

The reactivity of three- and four-membered heterocycles is dominated by ring strain, which facilitates ring opening. Four-membered rings often cleave to two two-membered fragments. Cationic three- and four-membered rings undergo nucleophilic attack when undergoing ring opening. Propiolactone undergoes ring opening, with the product being dependent on nucleophile and acidity of the reaction medium.



ring opening reactions

Anions of small heterocycles are uncommon. Extrusion of small gaseous fragments (e.g., N_2, CO_2) is very common for small, neutral heterocycles. *Cis–trans* isomerizations via 1,3-diradicals are often observed when three-membered heterocycles are heated or irradiated.

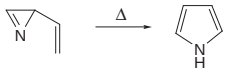


cis-trans isomerization of aziridines

Ring valence isomerizations of small rings are known. Ring expansion of vinyl heterocycles is well known. In

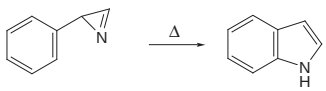
valence isomerization of an aziridine

many reactions of this form, especially when a three-membered heterocycle reacts *inter* molecularly with an olefin, the heterocycle acts as a 1,3-dipole, by a formal ring opening. Other ring expansions include the formation of



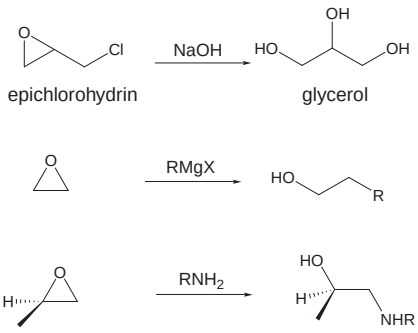
formation of pyrrole by ring expansion

indole from 3-phenyl-1-azirine. Such ring expansions are typical of three- and four-membered heterocycles, chiefly because of relief of ring strain.



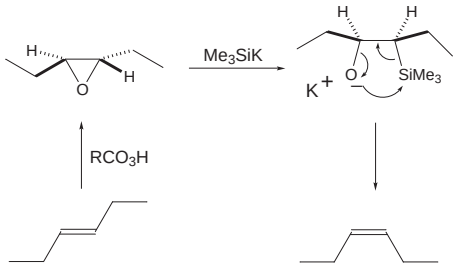
3-phenyl-1-azirine indole

Epoxides (oxiranes) are frequently used as synthetic intermediates and are usually prepared from alkenes by treatment with peracids or metal–hydroperoxide systems (see Section V.A.1 above). Epichlorohydrin has been used to prepare glycerol commercially. The strain of the small ring leads to ready attack of epoxides by a wide variety of nucleophiles.

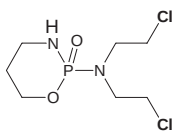


epichlorohydrin glycerol

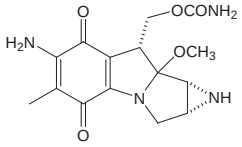
An useful procedure for inverting of alkene geometry involves stereospecific epoxidation and ring opening.



trans-3-hexene *cis*-3-hexene

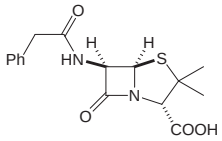


cyclophosphamide



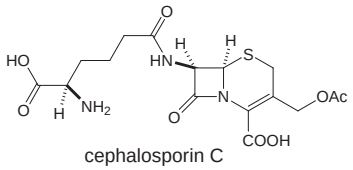
mitomycin C

Azetidine, the four-membered analog of aziridine, is a colorless liquid, stable at room temperature and less reactive than aziridine. Interest in the cyclic amide azetidinone rose rapidly in 1943, when it was correctly suggested that penicillins contained this heterocyclic ring. The fused azetidinone ring is the key to the reactivity of the two related series of antibioties, the penicillins and the cephalosporins.



benzylpenicillin

Several of the early chemotherapeutic agents used in a wide variety of cancer treatments depend upon an three-membered ring undergoing nucleophilic attack by an oxygen or nitrogen atom in DNA of the cancer cells. Examples include cyclophosphamide (which can eject chloride with the formation of an aziridinium ring, a powerful electrophile) and mitomycin. In contrast, epoxides derived by metabolism of many polybenzenoid hydrocarbons are highly toxic because the epoxides form covalent linkages with normal cell DNA.



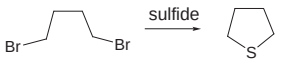
cephalosporin C

B. Five-Membered Heterocycles with One Heteroatom

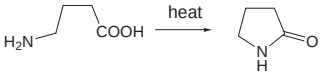
1. Synthesis

Pyrrole, furan, and thiophene are the most important parent heterocycles of this class, and their annelated derivatives constitute a vast array of both naturally occurring and synthetic compounds.

The ring closure of a four-carbon chain with the incorporation of one heteroatom is very easy.

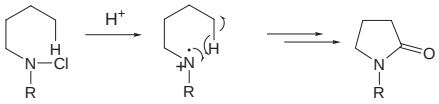


tetrahydrothiophene



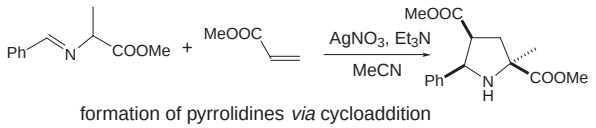
pyrrolidone

A special case of this is the Hofmann–Loeffler–Freytag reaction, in which pyrrolidines (or piperidines) are formed by the decomposition of protonated *N*-haloamines.



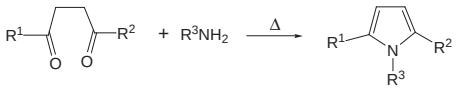
conversion of *N*-chloramines into pyrrolidines

This particular free-radical mechanism confers high selectivity in the formation of products owing to a preferred six-membered transition state. Another common route to five-membered heterocycles is via 1,3-dipolar cycloaddition.



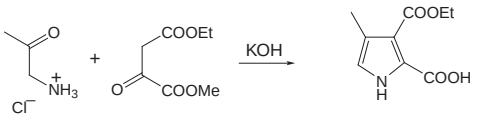
formation of pyrrolidines *via* cycloaddition

Fully unsaturated five-membered heterocyclic rings are often synthesized by the ring closure of 1,4-dicarbonyl compounds to give, with ammonia or primary amines, pyrroles (Paal–Knorr synthesis); with dehydrating agents, furan; and with phosphorus pentasulfide, thiophenes.



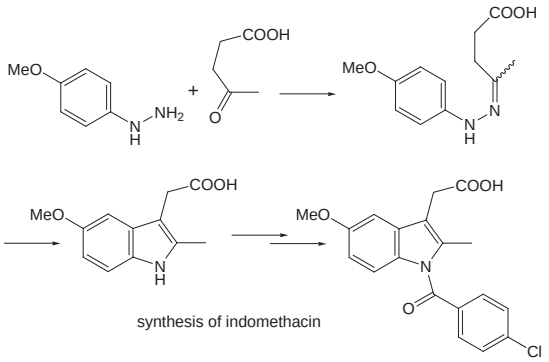
Paal-Knorr synthesis of pyrroles

Another important synthesis of pyrroles is the Knorr pyrrole synthesis, in which alpha-aminoketones or their equivalent are condensed with carbonyl compounds containing active alphamethylene groups.



Knorr pyrrole synthesis

Indoles (benzo[*b*]pyrroles) are frequently prepared by the Fischer indole synthesis, in which aryl hydrazones undergo rearrangement on heating, usually in the presence of acid. The important drug indomethacin, used to combat rheumatoid arthritis, has been synthesized by this method.

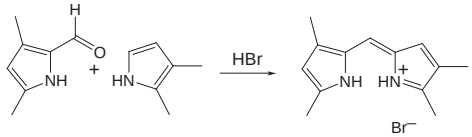


synthesis of indomethacin

2. Naturally Occurring Five-Membered Heterocycles

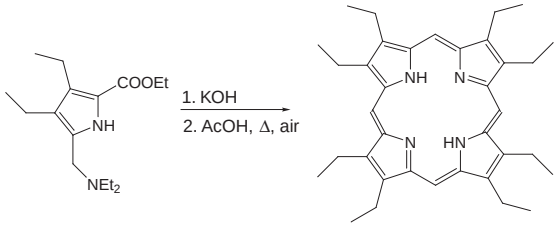
Pyrrole (*Gk·pyr* = fire) was first isolated from bone pyrolyzate; it imparts a bright red color to pinewood moistened with concentrated hydrochloric acid.

Condensation of suitable pyrroles in the presence of acid leads to dipyrromethene cations, which are precursors to porphyrins, a group of compounds found in all living matter and forming the basis of respiratory pigments.



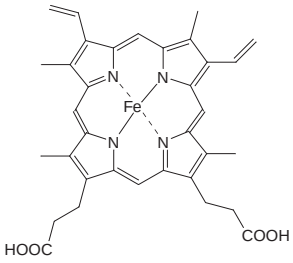
a dipyrromethene cation

The oxidative condensation of certain pyrroles can lead directly to the porphyrin nucleus, as in the case of octaethylporphyrin.

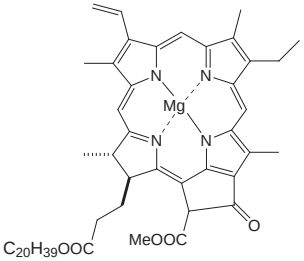


octaethylporphyrin

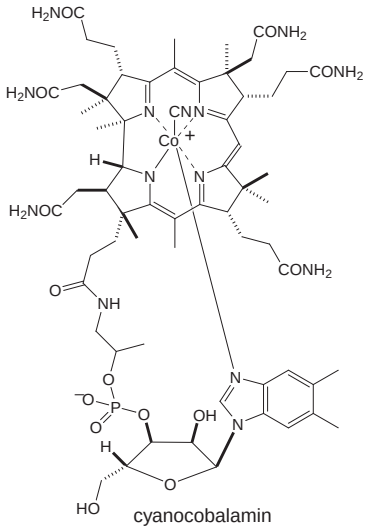
Heme, the coloring portion of hemoglobin, and the chlorophylls, the green pigments of plants, are both porphyrins.



heme



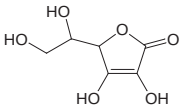
chlorophyll a



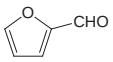
cyanocobalamin

Vitamin B_{12} (cyanocobalamin) is used to combat pernicious anemia; it was synthesized by Woodward and Eschenmoser in brilliant and unparalleled sequences of over 30 reactions. The skeleton of vitamin B_{12} is the corrin macrocycle, differing from a porphyrin ring by having one less carbon atom in the linking of the four pyrrole rings. Here, the nitrogen atoms of the pyrrole rings bind (act as ligands to) a central carbon atom. Moreover, a nitrogen atom of a benzimidazole (heterocyclic) ring also acts as a ligand. In this biologically active and important molecule, it is unsurprising to find a further type of heterocycle: a ribose ring, one of the sugars based on a hydrogenated furan ring.

Other examples of naturally occurring furans are vitamin C (ascorbic acid), first isolated from orange juice in 1928, and furfural, the artificial oil of ants, from which the solvent tetrahydrofuran is made.

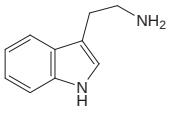


ascorbic acid

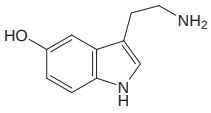


furfural

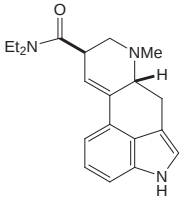
Many important indoles are found in nature: tryptamine, the parent of the essential amino acid tryptophan; the powerful vasoconstrictor serotonin; and the powerful hallucinogen lysergic acid diethylamide (LSD) are a few examples.



tryptamine

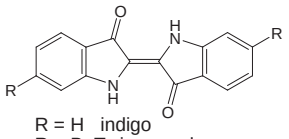


serotonin



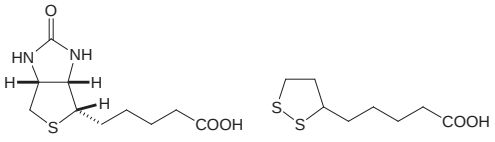
lysergic acid diethylamide

Indoles have been used since ancient times; woad, the coloring matter worn by Boudicca's warriors, was an impure form of indigotin (indigo) extracted from the plant *Indigofera tinctoria*. Tyrian purple, the highly prized dye worn on the garments of emperors, is 6,6′-dibromoindigo, and was extracted from the mollusc murex.



R = H indigo
R = Br Tyrian purple

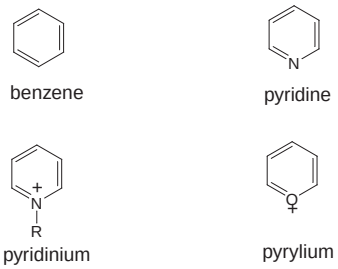
Thiophene itself is found in coal tar. Biotin, the human growth factor vitamin H, contains a tetrahydrothiophene ring alpha-Lipoic acid, when enzyme-bound, is converted into acetyl lipoate, which acetylates coenzyme A giving acetyl coenzyme A, the biological building block that supplies two carbon atoms.



biotin alpha-lipoic acid

C. Six-Membered Rings with One Heteroatom

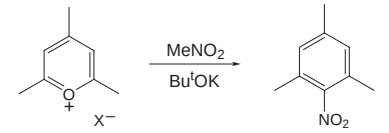
Benzene, pyridine, pyridinium, and pyrylium cations form an isoelectronic series with decreasing electron densities at ring carbon atoms and have increasing susceptibility to nucleophilic attack.

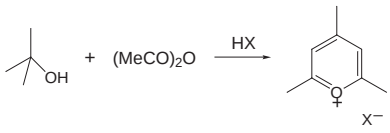


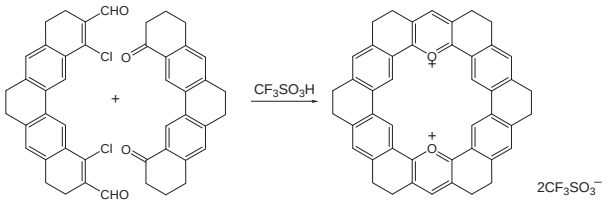
benzene pyridine pyridinium pyrylium

Pyrylium salts are usually synthesized by condensation methods.

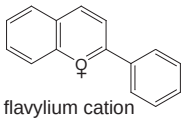
Since they are very reactive, pyrylium salts undergo many trasnformations, giving pyridines with ammonia, pyridinium salts with primary amines, and even benzenes with certain carbon nucleophiles, all by S_N ANRORC mechanisms (nucleophilic substitution with addition of nucleophile, ring opening, and ring closure).





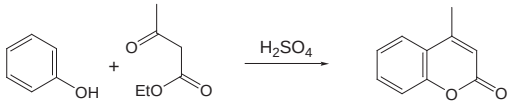


Benzannelated pyrylium salts are naturally abundant; anthocyanin pigments of red and blue flowers are a combination of sugars with anthocyanidins, of which flavylium is the parent cation.



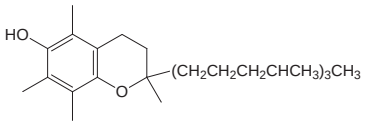
flavylium cation

Pyrones, pyrans, and their benzannelated derivatives can be obtained from the corresponding pyrylium salts. Diels–Alder reactions are also useful routes to pyrans. Coumarins (benzo[*b*]pyrones) are often prepared by the von

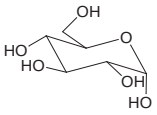


synthesis of 4-methylcoumarin

Pechmann reaction, in which phenols are condensed with beta-ketoesters in the presence of a catalyst. Coumarin itself is a flavor additive found in sweet clover. Vitamin E (alpha-tocopherol) possesses properties of antisterility and is a fused pyran. In the body, vitamin E acts as an antioxidant which after oxidation can be regenerated by vitamin C (ascorbic acid).



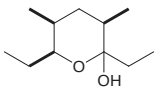
vitamin E



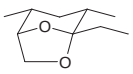
alpha-D-glucose

The hydroxypyran alpha-D-glucose is a main source of energy for living organisms. Cellulose and starch are polymers of glucose.

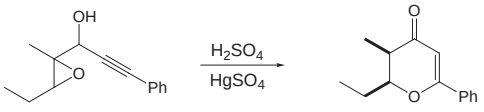
Saturated or partially saturated pyran rings are constituents of many pheromones such as serricornin, a principal pheromone attractant secreted by the cigarette beetle, a major pest of cured tobacco leaves, and (−)-alpha-multistriatin, an important component of the pheromone bouquet of the elm bark beetle. One route that allows highly stereocontrolled placement of substitutents involves a mercury(II)-catalyzed rearrangement of an acetylenic epoxy alcohol.



serricornin



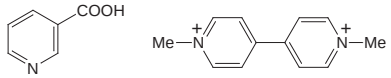
multistriatin



1. Pyridines and Pyridinium Salts

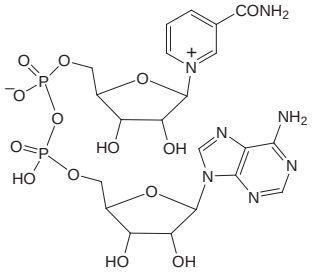
A simple route to pyridinium salts is by direct alkylation of pyridines, the Menschutkin reaction.

The pellagra-preventive nicotinic acid and the powerful weedkiller paraquat are examples of important pyridine derivatives.



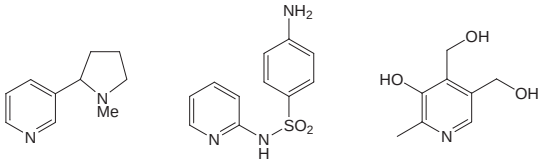
nicotinic acid paraquat

The antagonist nicotinamide adenine dinucleotide (NAD^+) also contains a pyridinium ring.



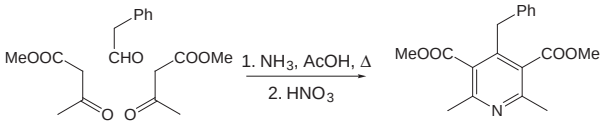
nicotinamide adenine dinucleotide (NAD^+)

Pyridine is present in bone pyrolyzates and coal tar, from which it was isolated in 1846 by Anderson. It is now prepared industrially from tetrahydrofurfuryl alcohol and ammonia. Nicotine (in tobacco), the antibacterial agent sulfapyridine, and the benzene cofactor pyridoxine (vitamin B_6) are three important pyridines.

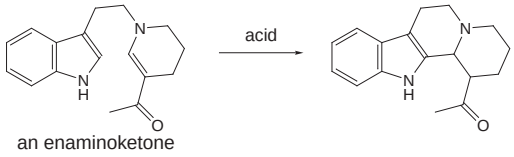


nicotine sulfapyridine pyridoxine

As for five-membered rings, common routes to six-membered rings involve condensation, ring closure, and cycloaddition reactions. In the Hantzsch pyridine synthesis, 2 moles of a beta-dicarbonyl compound are condensed with 1 mole of an aldehyde in the presence of ammonia to give, after oxidation, a pyridine.



Hydrogenated derivatives of pyridine are often used in alkaloid synthesis; heterocyclic enamines undergo various cyclization and annelation reactions. Complete hydrogenation of pyridine leads to piperidine. Coniine, or alpha-propylpiperidine, is a poisonous principle in hemlock, from which Socrates died. The synthesis of coniine by Ladenburg in 1886 is the earliest recorded synthesis of an alkaloid.

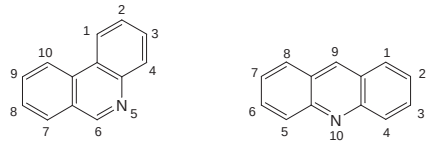


an enaminoketone

piperidine coniine

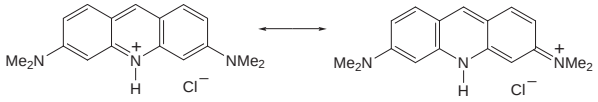
2. Quinolines, Isoquinolines, and Their Derivatives

Formal benzannelation of pyridine gives quinoline and isoquinoline; bisbenzannelation gives phenanthridine and acridine. Whereas the former is numbered systematically, it should be noted that acridine is not, the central ring being numbered finally.



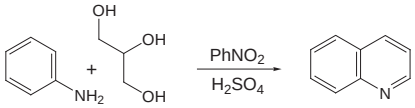
phenanthridine acridine

Acridine dyestuffs, such as acridine orange L, owe their color to the extensive delocalization of cationic charge.



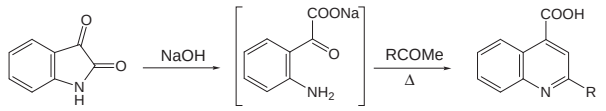
acridine orange L

There are several important routes to quinolines. In the Skraup reaction, an aniline is heated with a mixture of glycerol, nitrobenzene, and sulfuric acid to give a quinoline.

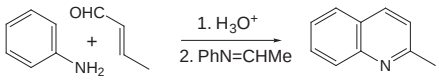


Skraup synthesis of quinoline

In the Pfitzinger reaction, isatic acids are condensed with alpha-methylene carbonyl compounds.

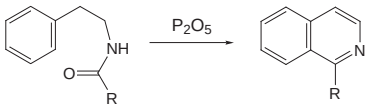


In the Doebner–Miller reaction, quinolines are prepared by condensing anilines with unsaturated aldehydes or their equivalent.



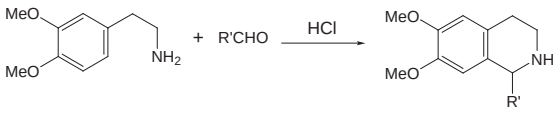
Doebner-Miller synthesis of 2-methylquinoline

Isoquinolines, abundant in nature, can be prepared by the cyclodehydration of beta-phenethylamides, a procedure known as the Bischler–Napieralski reaction.



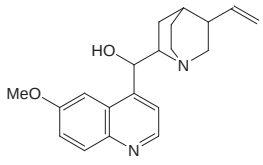
Bischler-Napieralski synthesis of dihydroisoquinolines

Isoquinolines are also formed in the Pictet–Spengler condensation of beta-arylethylamines with carbonyl compounds.



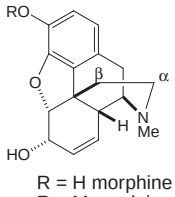
Pictet-Spengler synthesis of tetrahydroisoquinolines

Quinoline and isoquinoline alkaloids constitute a vast and important array of nitrogenous bases. The antimalarial quinine, a major alkaloid found in cinchona bark, contains a methoxy-substituted quinoline ring.

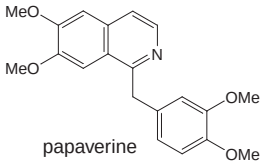


quinine

Morphine, codeine, and papaverine (*L. papaver* = poppy) are all found in opium; they are all isoquinoline alkaloids. Morphine and codeine are well-known analgesics, while papaverine is a potent coronary and cerebal vasodilator.

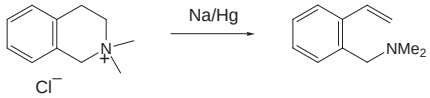


R = H morphine
R = Me codeine



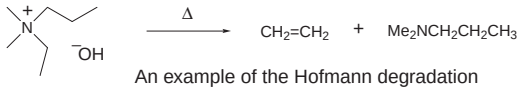
papaverine

The Emde and Hofmann degradations have been used in the structural elucidation of numerous naturally occurring substances, including many alkaloids. In the Emde degradation, quaternary ammonium salts are reductively cleaved by sodium amalgam, *N*-allyl and *N*-benzyl groups being preferentially cleaved.



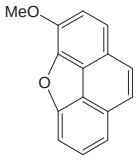
Emde degradation of a quaternary ammonium salt

In the Hofmann degradation, pyrolysis of a quaternary ammonium hydroxide induces beta-elimination. Often, an unknown base is alkylated with methyl iodide, then treated with base, and the procedure is repeated exhaustively. The fragments so obtained often help to elucidate the structure of the nitrogen base.



An example of the Hofmann degradation

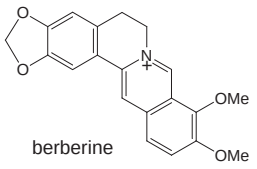
When the Hofmann degradation is applied to codeine, by quaternization with dimethyl sulfate and subsequent elimination of the quaternary salt with base, methylmorphenol is formed.



methylmorphenol

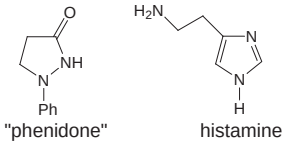
The power of this degradative method is shown by the formation of the relatively simple and identifiable four-ring skeleton from the complex structure of codeine. In early work, the structures of many alkaloids were elucidated by a combination of degradative and synthetic methods.

Berberine is an antimalarial isoquinolinium salt that was first isolated from *Berberis vulgaris* in 1837.

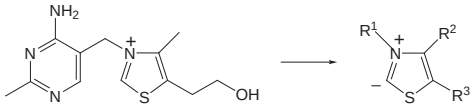


D. Five- and Six-Membered Rings with Two or More Heteroatoms

An important pyrazolone is the developing agent "phenidone," widely used in black-and-white photography. Histamine is an imidazole connected with many

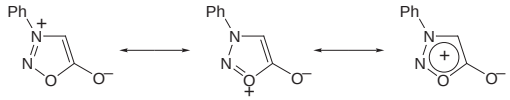


pathological conditions, including allergies—hence the development of antihistamine drugs. Derivatives of the biologically important thiamine (vitamin B_1) assist in decarboxylation and transketolase reactions by forming a nucleophilic zwitterion based on the thiazolium ring.



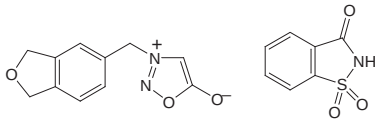
conversion of thiamine into a biologically active zwitterion

The sydnones are aromatic heterocyclic betaines that have been called "mesoionic." They have found use as antimalarial agents.

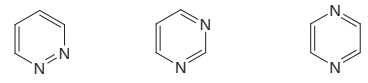


canonical forms of a sydnone

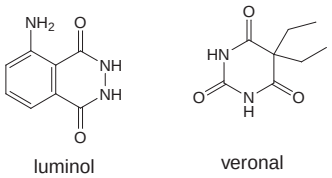
A well-known derivative of isothiazole is saccharin, 500 times sweeter than sugar.



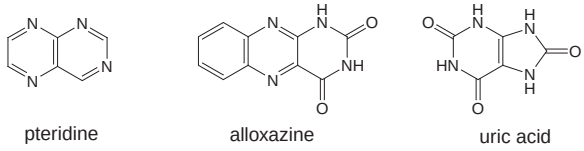
The three parent monocyclic diazines are pyridazine, pyrimidine, and pyrazine.



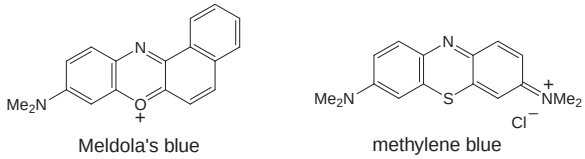
Luminol is a pyridazine that upon oxidation gives a striking emission of light. Veronal is a synthetic barbiturate that acts as a sedative and hypnotic. However, the



pyrimidines are the most important class of compounds in this group; the nucleic acids (see next subsection) are based on the pyrimidine ring. Other important nitrogen heterocycles whose derivatives occur in nature are pteridine, alloxazine, and uric acid (a metabolic product).

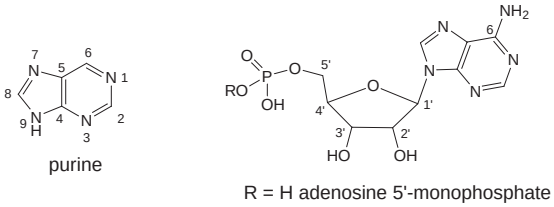


There are many dyes based on cationic six-membered heterocyclic rings, such as Meldola's blue and methylene blue, a biological staning agent.



1. Nucleic Acids

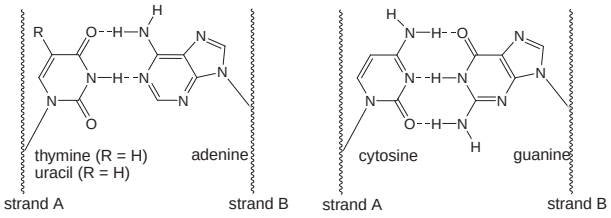
Nucleotides comprise a D-ribose or 2-deoxy-D-ribose sugar, which is linked by C-5′ to a phosphate unit and by C-1′ to a nitrogen of one of six heterocycles, of which three are pyrimidines—uracil, cytosine, and thymine—and three are purines—adenine, guanine, and hypoxanthine.



Adenosine 5′-monophosphate is a typical nucleotide; it contains the heterocyclic adenine 6-aminopurine. The di- and triphosphates (ADP and ATP) play central roles in biological phosphorylation.

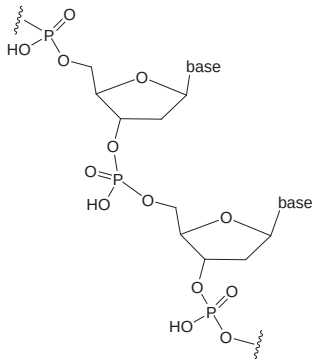
Nucleic acids are polynucleotides. They are of two main types: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Both are present in every living cell; they direct the synthesis of proteins and are responsible for the transfer of genetic information.

In 1953, Watson and Crick postulated that DNA was formed by the twisting of two polynucleotide chains into a right-handed double helix. The chains are held together by the hydrogen bonding of two base pairs: thymine– or uracil–adenine and cytosine–guanine. No other combination of those base pairs is possible.



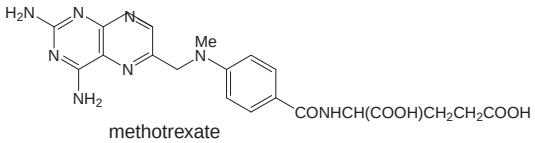
combination of base pairs in nucleic acids

A segment of a DNA phosphodiester chain with the positions of attachment of the heterocyclic bases is shown.

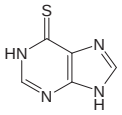


segment of a DNA phosphodiester chain

5-Fluorouracil, used to treat breast and stomach cancers, resembles uracil, without which the cell cannot make pyrimidine nucleotides. 5-Fluorouracil enters the biosynthetic chain but forms an inactive complex which blocks the synthesis of pyrimidine nucleotides and hence inhibits DNA synthesis in cancer cells. Methotrexate is a derivative of folic acid and interferes with folic acid metabolism by inhibiting the enzyme dihydrofolate reductase, essential for the synthesis of purine and pyrimidine, and therefore essential to the synthesis of DNA. Methotrexate finds use in treating childhood leukemia and other cancers, particularly in combination with other drugs. 6-Mercaptopurine blocks purine nucleotide synthesis and is a purine antimetabolite in clinical use.



methotrexate

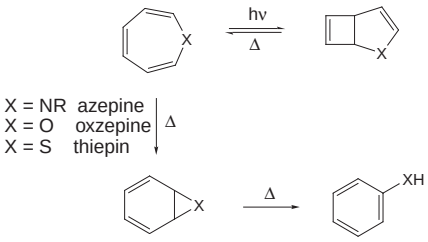


6-mercaptopurine

E. Rings with Seven or More Members

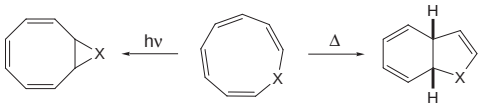
1. Synthesis

The famous rules of Woodward and Hofmann govern, among many other reactions, the contractions and expansions of heterocyclic rings with seven or more members. A seven-membered heterocyclic ring can give a 4,5-system on photolysis, as well as a benzene ring on thermolysis.



transformations of seven-membered heterocycles

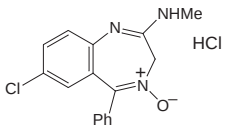
A nine-membered heterocyclic ring can give a 3,8-ring system on photolysis, and also a 5,6-ring system on thermolysis. Some of these reactions are of considerable synthetic importance.



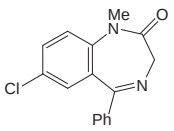
transformations of nine-membered heterocycles, X = O, NR

2. Properties

Azepine, oxepine, and thiepin have eight pielectrons (including the lone pair on the heteroatom). Thus, they have little aromatic character and are very reactive. The tranquilizing properties of certain diazepines, such as Librium and Valium, have prompted much research in this area.

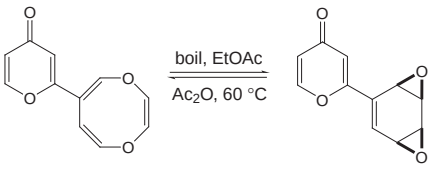


librium



valium (diazepam)

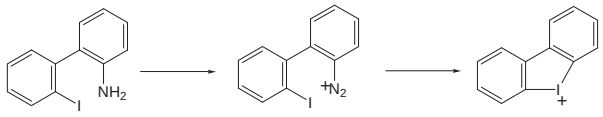
Many heterocycles possessing rings with more than seven atoms are known, some of which are natural products. A recently discovered antibiotic, LL-Z1220, undergoes reversible valence isomerization to a 1,4-dioxocin, an eight-membered heterocycle.



interconversion of 1,4-dioxocin and antibiotic LL-Z1220

F. Rings with Uncommon Heteroatoms

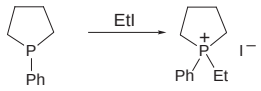
Nitrogen, oxygen, and sulfur are common, but by no means the only known heteroatoms in heterocyclic rings. For example, iodonium salts can be prepared by treating aminobiphenyls with nitrous acid, followed by ring closure of the diazonium salt.



iodonium salts *via* diazotization

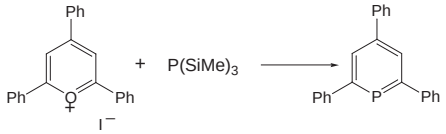
The corresponding chloronium and bromonium ions are not very stable.

Phosphorus, arsenic, and antimony form heterocyclic compounds fairly readily, especially the first two. 1-Phenylphosphonane was prepared as early as 1916, and with ethyl iodide gives a quaternary salt, just as do nitrogen heterocycles.



quaternization of 1-phenylphospholane

Phosphorins, in which phosphorus replaces one carbon atom in a benzene ring, can be prepared from pyrylium salts.

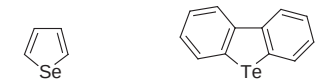


preparation of 2,4,6-triphenylphosphorin

Examples of arsenic and antimony heterocycles are isoarsindoline and stibabenzene.

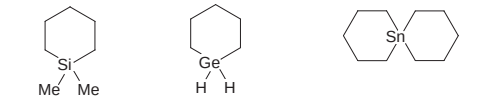
isoarsindoline stibabenzene

Like sulfur, selenium and, less commonly, tellurium form heterocyclic rings, such as selenophene and dibenzotellurophene.



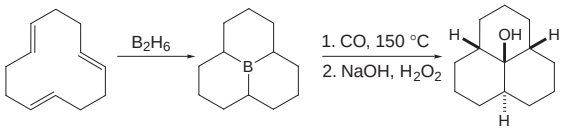
selenophene dibenzotellurophene

Silicon, tin, and germanium, although in the same periodic group as carbon, do not form stable double bonds; consequently, their heteroaromatic derivatives are unknown. However, many saturated heterocyclic derivatives are known, including 1,1-dimethylsilacyclohexane, germacyclohexane, and 6-stannaspiro[5,5]undecane.



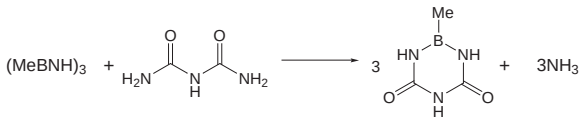
1,1-dimethyl silacyclohexane germacyclohexane 6-stannaspiro [5,5] undecane

Finally, many boron heterocycles have been reported. The procedure of hydroboration, chiefly for which H. C. Brown received the Nobel Prize, is illustrated below; subsequent treatment with carbon monoxide affords replacement of boron by carbon.



carbonylation of a boron heterocycle

A variety of boron heterocycles are known; one such is prepared by reacting a borazine with biuret.



preparation of a boron heterocycle

It can be expected that many new boraheterocycles will be prepared in the near future.

SEE ALSO THE FOLLOWING ARTICLES

ALKALOIDS • ANALYTICAL CHEMISTRY • BIOPOLYMERS • ELEMENTAL ANALYSIS, ORGANIC COMPOUNDS • ORGANIC CHEMICAL SYSTEMS, THEORY • ORGANIC CHEMISTRY, COMPOUND DETECTION • ORGANIC CHEMISTRY, SYNTHESIS • ORGANIC MACROCYCLES

BIBLIOGRAPHY

Acheson, R. M. (1976). "An Introduction to the Chemistry of Heterocyclic Compounds," 3rd ed., Wiley, New York.

Coffey, S. (ed.), (1977). "Rodd's Chemistry of Carbon Compounds," Elsevier, New York.

Gilchrist, T. L. (1997). "Heterocyclic Chemistry," Longman, Harlow, U.K.

Joule, J. A., Mills, K., and Smith, G. F. (1995). "Heterocyclic Chemistry," Chapman and Hall, London.

Katritzky, A. R. (ed.) (1963–1972). "Physical Methods in Heterocyclic Chemistry," Academic Press, New York.

Katritzky. A. R. (ed.), (1963–2000). "Advances in Heterocyclic Chemistry," Academic Press, New York.

Katritzky, A. R., and Rees. C. W. (eds.). (1984). "Comprehensive Heterocyclic Chemistry," Pergamon Press, Oxford.

Katritzky, A. R., Rees, C. W., and Scriven, E. F. V. (1996). "Comprehensive Heterocyclic Chemistry II," Pergamon Press, Oxford.

Meyers. A. I. (1974). "Heterocycles in Organic Synthesis," Wiley, New York.

Pozharskii, A. F., Soldatenkov, A. T., and Katritzky, A. R. (1997). "Heterocycles in Life and Society," Wiley, New York.

Suschitzky, H., and Scriven, E. (1989). "Progress in Heterocyclic Chemistry," Pergamon Press, Oxford.

Van der Plas, H. C. (1973). "Ring Transformations of Heterocycles," Academic Press, New York.

Weissberger, A., and Taylor, E. C. (eds.). (1950–1996). "The Chemistry of Heterocyclic Compounds," Wiley, New York.

Organic Chemical Systems, Theory

Josef Michl
University of Colorado

GLOSSARY

Electron affinity Energy released when an electron is brought from infinity and added to a molecule (it may be negative if energy actually has to be provided).

Electronegativity Measure of the tendency of an atom or an orbital to attract and accommodate electron density.

Gradient of a surface The gradient at a point is a vector directed up the steepest slope of the surface at that point; its length is a measure of the slope.

Improper rotational axis of symmetry (of order *n*) A molecule is said to possess such an axis located in a particular direction if rotation by an angle $2\pi/n$ about that direction, followed by mirroring in a plane perpendicular to the direction, converts the molecule back to itself.

Inner-shell electrons Electrons in orbitals of energy lower than that of the valence shell (closer to the nucleus).

Interaction matrix element The interaction matrix element of the Hamiltonian H between wave functions ϕ_1 and ϕ_2 is given by $\int \phi_1 H \phi_2 \, d\tau$, where integration is over all space.

Ionization potential Minimum energy that must be provided to a molecule in order to remove one of its electrons to infinity.

Ortho, meta, para Designation of relative positions of two substituents on a benzene ring: located on adjacent ring carbons (ortho), on next-nearest-neighbor ring carbons (meta), and on carbons across the ring from one another (para).

Pauli principle A wave function describing the state of a system containing two or more electrons is antisymmetric with respect to the exchange of all coordinates of any two electrons (i.e., is converted to minus itself on such an exchange). One of the consequences is that two electrons of the same spin cannot reside in the same orbital.

THE THEORY of organic chemistry deals with the fundamental concepts that underlie and unify the experimental

observations made by chemists working with organic molecules. It is believed that the behavior of molecules can be understood, in principle, in terms of a few basic laws of physics and that it is only the mathematical complexity of the resulting equations that limits the accuracy with which the behavior of organic molecules can be predicted *a priori*. Despite the largely approximate nature of the theoretical treatments applicable to large molecules, the theory has made substantial contributions, primarily by providing the language through which the various observed phenomena can be interrelated. It permits the rationalization of trends and at times of individual observations concerning the reactivity and properties of organic molecules, and in some instances it has provided useful predictions.

I. CLASSICAL BONDING THEORY

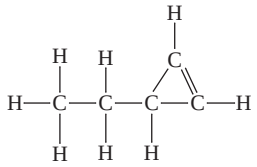
Although the theory of organic chemistry has now been cast in terms of quantum theory, most of the older qualitative concepts of the classical bonding theory remain useful. The classical theory thus represents a suitable introductory level for the subject at hand.

A. Structural Formulas

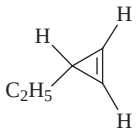
In the classical description an organic molecule is represented by a structural formula. This is a collection of atomic symbols (C, H, O, N, etc.) representing atoms (including their inner-shell electrons but excluding their valence-shell electrons; the *d* electrons of a transition metal atom are included, although they may participate in bonding). At least one, but usually many, of these atoms must be carbon in order for the molecule to qualify as organic. The atomic symbols are connected by a network of single, double, and triple lines, which stand for single, double, and triple covalent bonds, respectively. These are formed by the sharing of electron pairs between atoms, with one bond representing one pair. The number of nearest neighbors to which an atom is attached is called its coordination number.

In addition, short lines (see **2a**) or pairs of dots can be used to represent unshared (lone) electron pairs on an atom, but these are frequently omitted. Single dots represent odd (unpaired) electrons if such are present (see **2a**–**2c**).

The symbols C and H for the carbon atom and the hydrogens attached to it, respectively, are also frequently omitted. Commonly occurring groups of atoms of well-known internal structure are often indicated by giving the kind and number of atoms involved (e.g., C_2H_5 for ethyl) or by an abbreviation (in this case, Et). A few examples are given in **1a** through **2c**:

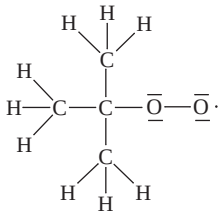


1a

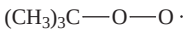


1b

1c



2a



2b

2c

The number of bonds formed by an atom (its "covalency") is dictated by the rules of valence. These state that in order for an organic molecule to have reasonable stability under ordinary conditions rather than to appear only as a transient reaction intermediate, if at all, the valence shells of all atoms in the structure have to contain a certain number of electrons: 2 for hydrogen, 8 for other main-group elements, and 18 for transition metal elements. The group of 8 electrons in the valence shell of an atom is often referred to as a valence octet. In order to determine the number of electrons in the valence shell of an atom, one counts all the unpaired electrons or electrons present in lone pairs on that atom, plus two electrons for each single bond in which the atom is participating (four for a double bond, six for a triple bond). In structures **1a** through **1c** all atoms satisfy the rules of valence; in structures **2a** through **2c** the terminal oxygen atom does not.

Each type of bond is associated with a contribution to the total energy of the molecule, and these contributions are approximately additive. Typical bond strengths are presented in Table I. These are to be taken only as a rough guide since the immediate environment of the bond, steric strain (Section I.B), and resonance (see Section I.C) can have significant effects.

Atoms with valence shells that contain fewer electrons than demanded by the valence rules are said to be coordinatively unsaturated and usually are carriers of high chemical reactivity (terminal oxygen in **2**, the central carbon in **3**). Atoms with valence shells that contain a larger number of electrons than dictated by the rules are said to be hypervalent. This situation is rare for atoms of the elements of the second row of the periodic table (presumably due to their small size and the resulting steric crowding) but fairly common for those of the third and lower rows, where the number of valence-shell electrons can be 10, 12, or even higher. Molecules containing hypervalent atoms are often stable, particularly if the hypervalent atom is of lower electronegativity than its neighbors (e.g., the tin

TABLE I Typical Bond Energies in Organic Molecules[a]

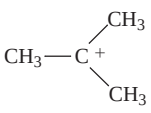
X	X—H	X—C	X—N	X—O	X═C	X≡C
C	100	81	69	84	148	194
N	93	69	38	43	148	213
O	110	84	43	33	172	
F	135	105	65	50		
Si	72	69		103		
P	77	63				
S	83	65			128	
Cl	103	79	48	50		
Br	88	67		53		
I	71	57		57		
N—N	38	N═N	100	N≡N	226	
O—O	33	N═O	145			
S—S	54	O═O	96			

[a] In kilocalories per mole; 1 kcal = 4.184 kJ.

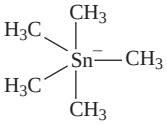
atom in **4**; the electronegativity of an element increases as one moves up and to the right in the periodic table).

Those molecules that contain an odd number of electrons cannot satisfy the rules for all of their atoms and are known as free radicals (e.g., **2**).

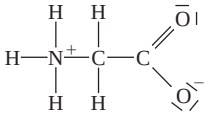
In addition to atomic symbols and symbols for bonds, lone pairs, and unpaired electrons, the classical structural formulas of organic chemistry also indicate atomic charges (e.g., **3**–**5**):



3



4



5

The way to determine the charge on an atom is to count the valence-electron ownership of an atom in the molecule and compare it with the number of valence electrons on a neutral isolated atom of the same element. If the two agree, the formal charge is zero. If there is one more electron on the atom in the molecule than on an isolated atom, the formal charge is minus one and so on.

In order to determine the valence-electron ownership of an atom in a molecular structure, one counts all electrons indicated as lone pairs as well as unpaired electrons on the atom, plus one for each bond in which the atom participates. Thus, one assumes that the two electrons of a bond are shared equally between the two atoms that it joins. To indicate that this is unrealistic when the two atoms differ in their electronegativity, the charges are referred to as formal. For molecules containing transition metal elements, the individual formal charges are frequently not indicated at all.

The sum total of formal charges on atoms in a molecule is equal to its net charge, and this is always indicated. Negatively charged molecules are called anions, positively charged ones cations. The electrostatic force of attraction between two oppositely charged ions is sometimes referred to as an ionic bond.

Typical bonding situations in which atoms of elements that are most commonly found in organic molecules find themselves in molecular structures are listed in Table II. Analogous bonding situations are found throughout each column of the periodic table, except that atoms of second-row elements resist hypervalency.

The hydrogen atom does not suffer from steric constraints in its ordinary univalent state, in which it makes only one bond. It can enter into a special kind of weak hypervalent interaction known as the hydrogen bond, which attaches it to a lone-pair-carrying second atom. The hydrogen bond is indicated by a dotted line. As usual for hypervalent interactions, hydrogen bonding is particularly important if the neighbors of the hydrogen atom are highly electronegative.

B. Molecular Geometries

Classical structural formulas imply molecular geometries. These are determined by bond lengths, valence angles, and dihedral angles and describe the average nuclear positions when the molecule is at equilibrium.

In real molecules, at least some vibrational and internal rotational motion is always present. In many organic molecules, this can be neglected in the first approximation, and the molecules can be considered rigid. Some, particularly those lacking rings and multiple bonds, are definitely floppy at room temperature due to nearly free rotation around single bonds but can be viewed as rigid at sufficiently low temperatures.

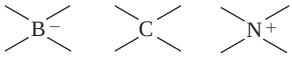
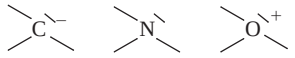
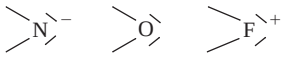
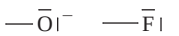
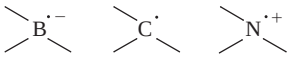
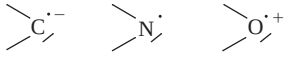
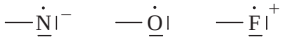
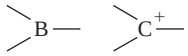
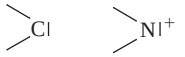
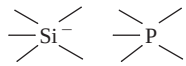
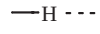
1. Bond Lengths

Bond lengths are generally determined by the nature of the two atoms bonded, with minor variations depending on the environment. Each kind of atom can be associated with the value of its "covalent radius." A bond length is approximately equal to the sum of the covalent radii of the participating atoms. Typical lengths of the most common bonds in organic molecules are listed in Table III.

2. Valence Angles

Valence angles are the angles between two bonds on the same atom, generally dictated by the coordination number of the atom. However, if lone pairs are present on an atom, each of these counts for yet another neighbor. On the other

TABLE II Common Building Blocks of Organic Molecules

Building blocks	Number of electrons in valence shell	Coordination number[a]
In agreement with the rules of valence		
	2	1
	8	4
	8	3
	8	2
	8	1
Coordinatively unsaturated		
	7	3
	7	2
	7	1
	6	3
	6	2
	6	1
Examples of hypervalent atoms		
	10	5
	10	4
	10	3
	(4)	(2)

[a] Parentheses are for hydrogen atom in a hydrogen bond, not always considered hypervalent.

hand, the presence of a single unpaired electron on an atom usually has only a minor influence on its valence angles.

For atoms forming two single or multiple bonds and carrying no lone pairs, the valence angles normally are in the vicinity of 180°; for atoms forming three single or multiple bonds and carrying no lone pairs they are ~120°; and for atoms forming two such bonds and carrying one lone pair, they are usually a little smaller. For atoms forming four bonds and carrying no lone pairs they are ~109°, and for atoms carrying a lone pair plus three bonds or two lone pairs plus two bonds they are a little smaller still.

The angles of 120° correspond to the center of an equilateral triangle being connected to its vertices, so that all three bonds are coplanar. The angles of 109° correspond to the center of a regular tetrahedron being connected to its vertices. The steric arrangement around an atom carrying a total of five bonds and lone pairs usually corresponds to a trigonal bipyramid and that around an atom carrying a total of six bonds and lone pairs to a regular octahedron, with the atom in the center in each case.

The valence angles given provide only a rough guide. Their exact values in any real molecule depend on the

TABLE III Typical Bond Lengths in Organic Molecules[a]

X	X—H	X—C	X═C	X≡C
B	1.21	1.56		
C	1.09	1.54	1.34	1.20
N	1.00	1.47	1.30	1.16
O	0.96	1.43	1.22	
F	0.92	1.38		
Si	1.48	1.84		
P	1.42	1.87		
S	1.34	1.81	1.56	
Cl	1.27	1.76		
Br	1.41	1.94		
I	1.61	2.14		
N—O	1.36			
N═O	1.21			
N—F	1.36			
N—Cl	1.75			

[a] In Å; 1 Å = 100 pm.

environment and, in particular, on the number of lone pairs on the atom.

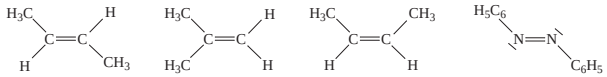
The bending motions are generally relatively easy at 180° valence angles, so that large excursions from the normally preferred angle are possible at room temperature. Similarly, out-of-plane vibrations around atoms characterized by 120° valence angles are also relatively easy, as is the interchange of positions of a lone pair and a single bond. Tetrahedral bond arrangements around an atom are relatively rigid. However, a flipping motion ("umbrella inversion") of a lone pair from one to the other side of the three bonds present is very facile on atoms of the second row (not those of lower rows). The interchange of ligand positions is easy on pentacoordinate atoms and difficult on hexacoordinate ones.

The presence of small rings in the molecule may introduce very large deviations from the usual valence angle values. For example, in cyclopropane the carbon atoms form an equilateral triangle so that the CCC angle is 60°, not much more than half of the normally expected value. Such deviations from the normally preferred angles are energetically unfavorable, and the molecule is said to exhibit angular strain.

3. Dihedral Angles

A dihedral angle is defined as the angle between two planes, both of which pass through the same bond. One of the planes also contains one of the additional bonds formed by one of the bond termini, and the other plane contains one of the additional bonds formed by the other terminus.

The preferred dihedral angles around a double bond are 0° and 180°, once again counting a lone pair on a terminus as another nearest neighbor. Thus, the usual geometries around C═C and N═N double bonds are planar (**6**–**9**):



6 **7** **8** **9**

However, small twisting distortions from planarity are relatively easy at room temperature.

There is a much weaker preference for particular values of the dihedral angle around single bonds, and rotation around such bonds is nearly free. Usually, the value of 0° ("eclipsed") is avoided, and values of around 60° ("staggered") to 90° are somewhat preferred, depending on the number of lone pairs on the termini.

The general rules just stated for bond lengths and angles permit the construction of mechanical molecular models either from balls and sticks or on a computer screen. The size of the balls that represent the volume of individual atoms is given by their van der Waals radii. The sum of the van der Waals radii of two atoms represents the distance of most favorable approach of these two atoms if they are not mutually bonded (e.g., atoms on neighboring molecules in a crystal). Values for these quantities are compiled in Table IV. Molecules in which two or more atoms that are not bonded to one another and are located at distances shorter than the sum of the van der Waals radii are strained by steric crowding and are less stable than otherwise expected. Often, it is possible to avoid some of this unfavourable interaction by a distortion of the valence angles.

4. Molecular Mechanics

It is possible to augment the set of bond energies by a set of energy increments for deviations from optimum bond lengths, valence angles, dihedral angles, and van der Waals distances as a function of the magnitude of each and to compute the energy of a molecule as a function of its geometry within the framework of such a "springs and balls" model. Equilibrium geometry can then be found by energy

TABLE IV Atomic van der Waals Radii[a]

Atom	Radius (Å)	Atom	Radius (Å)	Atom	Radius (Å)
H	1.2	O	1.4	F	1.4
N	1.5	S	1.9	Cl	1.8
P	1.9	Se	2.0	Br	2.0
As	2.0	Te	2.2	I	2.2
Sb	2.2				

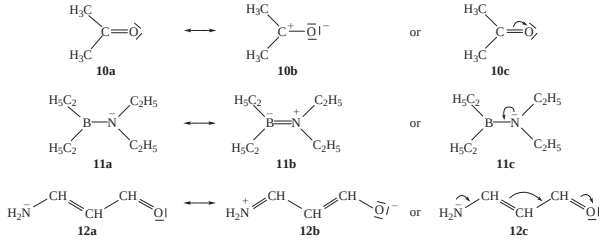
[a] 1 Å = 100 pm.

optimization. This approach is known as molecular mechanics and provides a good approximation, particularly for hydrocarbons, for which extensive and carefully optimized parameter sets are available. It runs into difficulties with molecules in which more than one classical valence structure is important.

C. Resonance (Mesomerism)

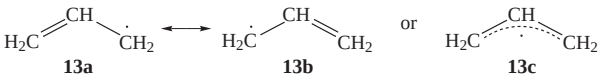
An important concept in classical structural theory is resonance (mesomerism). This is related to the fact that more than one classical structure can normally be written for a molecule. A double-headed arrow is usually placed between the structures to indicate that all of them contribute to the description of bonding in the molecule. It is important to note that all of the structures refer to the same molecular geometry.

The situation is most easily exemplified in the case of multiple bonds (e.g., **10**–**12**). The existence of two or more valence structures can also be indicated by curved arrows, as in **10c** to **12c**.



The relative importance of several contributing structures is dictated primarily by their energies (i.e., the energies of hypothetical molecules in which only the structure in question would contribute). The energies can be estimated qualitatively. Low energy is favored by the presence of a large number of bonds in the structure (a highly twisted double bond does not yield much stabilization—steric inhibition of resonance), by the absence of unfavorable charge separations, and by the presence of negative charges on electronegative atoms and positive charges on electropositive atoms.

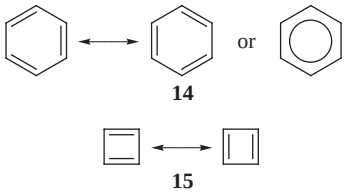
Two structures contribute equally when they are equivalent by symmetry, as in the allyl radical **13**:



A bond that is single in some and absent in other contributing resonance structures is called a partial bond and is often drawn with a dashed line.

The existence of several contributing structures is associated with a more or less significant effect on the thermodynamic stability of the molecule relative to that expected from the simple rules for any one of the contributing structures. Usually, it results in a stabilization, referred to as resonance energy.

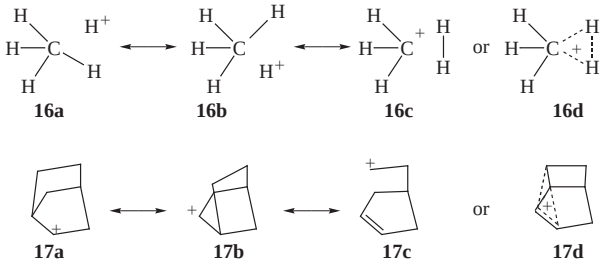
Resonance energy plays a particularly important role in cyclic systems. Those cyclic systems with two equivalent doubly bonded structures that contain $4N+2$ electrons in the perimeter (N is an integer) exhibit a large stabilization and are known as aromatic. The stabilization is known as the aromatic resonance energy. The archetypical example is benzene (**14**). Those containing $4N$ electrons in the perimeter are actually destabilized (antiaromatic); a good example is cyclobutadiene (**15**):



In the case of polycyclic ring systems of conjugated double bonds it is more difficult to specify the degree of aromatic stabilization or antiaromatic destabilization. A good rule of thumb is to count the number of possible structures, ignoring those with an even number of double bonds located inside any single ring. The higher the number, the larger the stabilization. These kinds of problems are far more efficiently handled by the more advanced quantum mechanical theories of molecular structure.

In order to simplify notation and to avoid writing a large number of contributing structures, it is customary to draw a circle inside an aromatic ring, as shown for **14**.

Ambiguities in the writing of molecular structures exist even in compounds containing only single bonds. In principle, it is possible to write the structures H^+H^-, H^-H^+, and H–H for molecular hydrogen. The charge-separated structures are ordinarily not written, and their existence is tacitly understood when H–H is written; this is true of all other bonds as well. In some cases the need to write more than one equivalent classical structure for a molecule containing only single bonds is not so easily avoided, and such structures are often called nonclassical (e.g., structures **16a**–**16d** for the CH_5^+ cation and structures **17a**–**17d** for the norbornyl cation):



Just as in the case of resonance involving double bonds, such systems with several equivalent structures are

frequently written with auxiliary symbols such as circles, often with dashed lines, particularly to indicate the delocalization of charge.

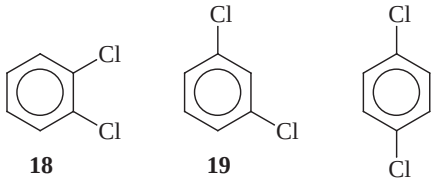
D. Isomerism

Molecules that do not differ in overall molecular formula (specified by the number of atoms of each kind and the net charge) but only in internal structure are called isomers.

1. Constitutional Isomers

Constitutional isomers are compounds that differ in connectivity, that is, in the way in which the constituent atoms are connected to one another. Graph theory is an important tool for their enumeration.

An important class of constitutional isomers are positional isomers, in which the functional groups are the same but differ in their location within the molecule (e.g., **6** and **7**, or the ortho, meta, and para isomers **18**–**20**):



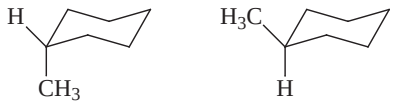
18 **19** **20**

2. Stereoisomers

Stereoisomers have the same connectivity but differ in the way in which the constituent atoms are oriented in space. They can be divided into configurational stereoisomers and conformational stereoisomers. The precise specification of the spatial arrangement of the groups in a configurational isomer is called its configuration, and in a conformational isomer, its conformation.

a. Configurational stereoisomers. These stereoisomers cannot be made superimposable by any rotations about single bonds. In order to make them superimposable, rotation about a double bond or a dissociation of one or more single bonds, or both, is necessary (e.g., **6** and **8**). Since these processes normally require considerable energy, they usually do not occur at a measurable rate at room temperature. Configurational stereoisomers can normally be isolated from one another and stored essentially indefinitely at room temperature.

b. Conformational stereoisomers. Often called conformers for short, these stereoisomers can be made superimposable by rotations about single bonds. Examples are the axial (**21**) and equatorial (**22**) conformers of a monosubstituted

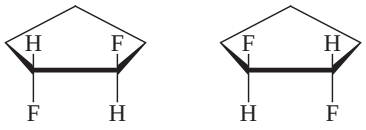


21 **22**

cyclohexane. Since rotations about single bonds are normally very facile, it is usually impossible to separate conformers from one another and to handle them separately at room temperature.

c. Chirality. Another important classification of stereoisomers into two groups is related to optical activity, manifested by the rotation of a plane of polarized light on passage through a sample.

A pair of stereoisomers that are related to one another in the same way as an object and its mirror image are called enantiomers (e.g., **23** and **24**):



23 **24**

Any pair of stereoisomers that are not related in this way are called diastereomers (e.g., **6** and **8** or **21** and **22**).

A molecule that is not identical with its mirror image is called chiral and occurs as a pair of enantiomers. The necessary and sufficient condition for chirality is the absence of an improper rotational axis of symmetry in the molecule (this includes a center of inversion and mirror plane symmetry elements).

A mixture of equal amounts of two enantiomers is known as a racemic modification and is optically inactive. A pure enantiomer or an unbalanced mixture of two enantiomers is optically active; the two enantiomers have opposite handedness and cause the plane of polarization to rotate in opposite directions.

II. QUALITATIVE MOLECULAR ORBITAL MODEL OF ELECTRONIC STRUCTURE

Although the classical structural theory as outlined so far accounts for much of organic chemistry, it leaves quite a few observations unexplained. In order to proceed and relate the chemical behavior of molecules to fundamental physical principles it is necessary to use quantum theory.

This can be done at a more or less rigorous quantitative level. A fairly direct translation of the structural concepts just outlined into mathematical terms then leads to the quantum mechanical valence–bond theory of electronic structure. Although conceptually appealing, computationally this method is quite unwieldy and has

not seen much use. An approach that is almost universally adopted nowadays is the quantum mechanical molecular orbital (MO) theory of electronic structure. This is computationally manageable and is described in some detail later.

Even the mathematically simpler MO approach to electronic structure requires the use of large computers for any quantitative applications. Although large-scale computations have been extremely valuable in enhancing the understanding of organic molecules, they are in themselves not appropriate for the day-to-day thinking of bench chemists. However, they have had a great influence on the much simpler qualitative models of electronic structure adopted for daily use by organic chemists and even some effect on that ill-defined body of knowledge generally referred to as the organic chemist's "intuition." The current qualitative model of molecular electronic structure based on the qualitative notions of MO theory represents a significant advance over the structural theory outlined in the preceding sections, but is not easy to describe unequivocally, since its form varies among individuals.

Current qualitative thinking about the electronic structure of organic molecules is based on the independent-particle model, in which it is assumed that the motion of each electron is dictated by the field of stationary nuclei and the time-averaged field of all the other electrons. In this model, any correlation of the instantaneous positions of the many electrons present is neglected.

Only in several well-recognized and more or less exceptional situations are correlation effects explicitly introduced. This is particularly true in the treatment of biradicals, which represent an important class of reaction intermediates but which are not discussed here, in the treatment of several other unusual bonding situations, and in the treatment of photochemical processes (e.g., in the consideration of differences in the reactivity of excited singlet and triplet states).

A. Atomic Orbitals

The independent-particle model is well known from the quantum mechanical description of atomic structure. Each electron in an atom is assumed to reside in an atomic orbital (AO) with a maximum of two electrons (of opposite spin) in any one orbital (Pauli principle). The AO is a function of the coordinates of one electron. The square of its magnitude at any point in space gives the probability density for finding the electron at that point. The magnitude itself can be positive or negative, with zero values at the boundaries of the positive and negative regions. The boundaries are referred to as nodal surfaces (planes, spheres, etc.). The energy of an electron residing in an orbital increases with the increasing number of nodal surfaces in the orbital.

Atomic orbitals are characterized by a principal quantum number n (there are $n-1$ radial nodes in an AO) and a letter indicating their shape as dictated by the angular nodes: An ns orbital has no angular nodes and is of spherical symmetry; each of the three equienergetic np orbitals has one angular mode (usually taken to be a plane through the nucleus, with respect to which the orbital is antisymmetric); each of the five nd orbitals has two angular nodes; and so on. Shorthand abbreviations for orbital shapes are shown in Fig. 1A. These are meant to indicate the regions of space in which the numerical value of the AO is the largest, also known as the lobes of an AO, as well as their signs.

In the ground state of an atom, the AOs are assumed to be occupied by the available electrons in the order of increasing energy of the subshells: $1s$, $2s$, $2p$, $3s$ . . . . The last at least partially occupied subshells and the more stable ones of the same principal quantum number represent the valence shell [in transition metals this contains the nd, $(n+1)s$, and $(n+1)p$ subshells]. Only valence-shell AOs are normally considered important for bonding: the $1s$ orbital in hydrogen and the $2s$ in lithium and beryllium ($2p$ are very close in energy and could be included as well), $2s$ and $2p$ in boron through neon, $3s$ in sodium and magnesium ($3p$ could be included), $3s$ and $3p$ in aluminum through argon, and so on.

The qualitative theory of bonding also makes use of combinations of valence AOs known as hybrid orbitals. Combining an s with a p orbital produces two equivalent sp hybrids pointing in opposite directions. Combining an s with two p orbitals produces three equivalent sp^2 hybrids pointing to the corners of an equilateral triangle. And combining an s with three p orbitals produces four equivalent sp^3 hybrids pointing to the corners of a regular

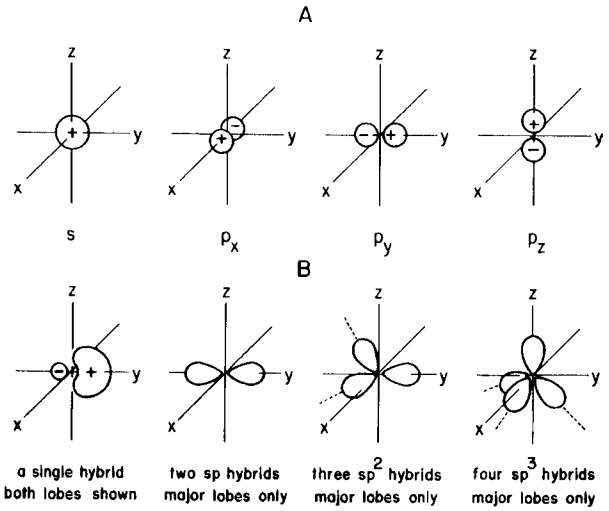


FIGURE 1 (A) Atomic orbitals. (B) Hybrid orbitals.

tetrahedron (Fig. 1B). Inequivalent hybrids can be used for other desired valence angles.

B. Molecular Orbitals

The independent-particle model can be extended to molecules by assuming that each electron again resides in an orbital of well-defined energy. Now, however, except for the inner-shell AOs, the orbital is spread over the space spanned by the molecular framework and is referred to as a molecular orbital. To a very good approximation, the inner-shell electrons behave in exactly the same way in the molecule as they would in an isolated atom, and we will not be concerned with them further. Instead, we concentrate on the valence MOs.

The delocalized form of MOs that we have just described is known as their canonical form. It can be shown that the wave function describing the ground electronic configuration of a molecule, containing a certain number of doubly occupied orbitals, does not change at all when these occupied MOs are mixed with one another in an arbitrary way. This degree of freedom in the wave function permits the construction of MOs that have been mixed in such a way that each one is localized in the smallest amount of space possible, using one of several possible criteria. The price one pays is that these localized MOs no longer have well-defined individual energies, although the total energy of the system is just as well defined as before.

It turns out that each of the localized MOs tends to be fairly well contained within the region of one bond or one lone pair; only in molecules with several important classical structures is this localization poor. On close inspection, the localization is actually never perfectly complete, and each localized orbital possesses weak "tails" extending to other parts of the molecule. Except for the exact nature of these tails, such a bond orbital often looks very much the same in all molecules that contain that particular bond, say, C–C. It is along these lines that one can begin to understand bond additivity properties and their failure in the case of molecules in which several classical resonance structures play an important role.

The standard qualitative model of molecular electronic structure requires the construction of a number of valence MOs sufficient to hold all the valence electrons. This is normally done separately for the electrons responsible for those bonds that are present in all important classical structures of the molecule ("localized" bonds, hence "localized" electrons, although strictly speaking, they are not really localized) and separately for the electrons responsible for partial bonds that are present in some and absent in other important classical structures ("delocalized" bonds, hence "delocalized" electrons).

The usual approach is based on the recognition of the fact that the formation of bonds in molecules represents only a small perturbation of electronic structure of the constituent atoms, bonding energies being of the order of 1% of total atomic energies. This is because the electric field in the vicinity of any one atom, where an electron spends most of its time, is dominated by the nucleus of this atom, since the force of attraction of an electron by a positive charge grows with the second inverse power of the distance from that positive charge.

This situation is acknowledged by assuming that each MO can be built by mixing AOs or hybrid orbitals.

C. Orbital Interactions

A brief consideration of the ways in which orbitals interact will be useful not only for a description of the procedure in which AOs are mixed to produce MOs, but also for subsequent reference to the mutual mixing of MOs, for instance, those of two molecules reacting with one another.

In the mixing of two orbitals, the important factors are their energies and the interaction matrix element (this is frequently referred to as the resonance integral β between the two orbitals). As a result of the mixing of n orbitals, n new orbitals result.

It is by far easiest to consider the case of only two mutually interacting orbitals ϕ_1 and ϕ_2 (Fig. 2). Let the energy of ϕ_1 be below that of ϕ_2. After the interaction, the energy

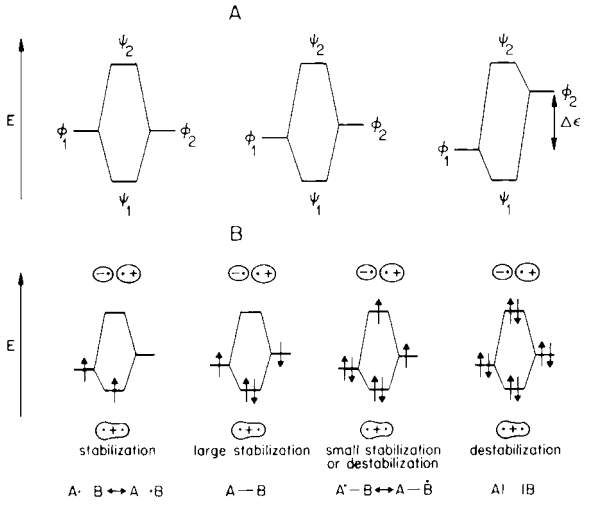


FIGURE 2 Orbital energies. (A) Interaction of orbitals ϕ_1 and ϕ_2 to produce linear combinations ψ_1 and ψ_2. (B) Interaction of atomic orbitals (outside levels) to produce a bonding and an antibonding combination (inside levels, orbital shape indicated). The effect on the net energy of the system is indicated for occupation with one to four electrons. Arrows represent electrons with spin up and spin down. At the bottom, the classical structural representation is shown for comparison.

of the more stable of the two new resulting orbitals ψ_1 will lie below that of ϕ_1 and that of the less stable one ψ_2 will lie above that of ϕ_2, as if the two original energies repelled one another. The amount by which each orbital energy is shifted will depend on the strength of the interaction element β and on the difference $\Delta\varepsilon$ in the original energies. If $\Delta\varepsilon \gg |\beta|$, the shifts are approximately inversely proportional to $\Delta\varepsilon$ (this approximation is known as first-order perturbation theory). When the two orbitals are originally degenerate ($\Delta\varepsilon = 0$), the shifts are the largest.

An important characteristic of the interaction between two orbitals ϕ_1 and ϕ_2 is their overlap integral $S_{12} = \int \phi_1\phi_2\, d\tau$, where the integration is over all space. For normalized orbitals ($S_{11} = S_{22} = 1$) the value of S_{12} can vary between -1 and $+1$. Usually if S_{12} is positive, the interaction element β is negative, and if S_{12} is negative, β is positive. If $S_{12} = 0$, orbitals ϕ_1 and ϕ_2 are said to be orthogonal; they can still have a nonvanishing resonance integral. The energy shifts caused by the interaction of two orthogonal orbitals are opposite in direction but equal in magnitude. When two nonorthogonal orbitals interact, the less stable new orbital ψ_2 is destabilized more than the more stable new orbital ψ_1 is stabilized.

The more stable of the two new orbitals is referred to as bonding and the less stable as antibonding with respect to the interaction considered. An orbital with the same energy as before the interaction is called nonbonding; such orbitals often result when more than two AOs are mixed. If two electrons are available to fill each bonding orbital, maximum stabilization will result relative to the situation before orbital interaction. If too many electrons are available and some must be placed into antibonding orbitals, some or all of the stabilization is lost and net destabilization may result. For this reason, the interaction of a doubly occupied with an unoccupied orbital leads to a net stabilization and the interaction of two doubly occupied orbitals to a net destabilization.

These general concepts can now be specialized to the case of mixing of AOs or hybrid orbitals to produce bond orbitals. Two valence-shell AOs or properly constructed hybrid orbitals located at the same atom are always orthogonal. For two orbitals located at the same center, the interaction element β is zero if one or both are pure AOs. However, two hybrid orbitals at the same center have a nonvanishing mutual interaction element. Its magnitude is related to the promotion energy (i.e., the energy difference between the AOs from which the hybrids were constructed).

Two AOs or hybrid orbitals located at different atoms can be, but do not need to be, orthogonal. A nonvanishing overlap integral between two p AOs can be produced in two geometrically distinct ways. When the axes of the two orbitals are aimed at one another, the overlap is said to be of the σ type; when they are parallel, it is said to be of the π type (Fig. 3). Intermediate situations are also possible. The interaction matrix element β between two AOs or hybrid orbitals located at different atoms is approximately proportional to the negative of their overlap integral S_{12}.

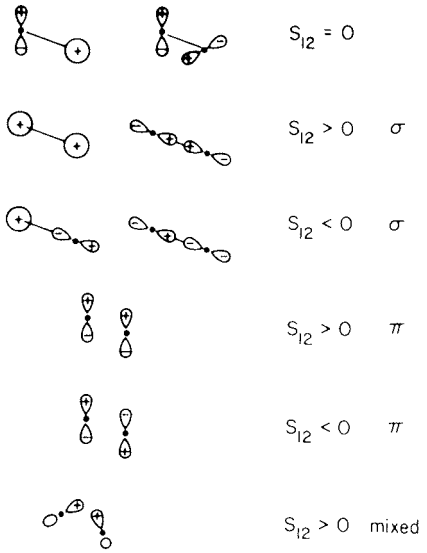


FIGURE 3 Overlap between orbitals.

The two new orbitals ψ_1 and ψ_2 that result from the interaction are linear combinations of the two old orbitals (Fig. 2). In the bonding orbital ψ_1, the overlapping portions of the entering old orbitals ψ_1 and ψ_2 have the same sign. In the antibonding orbital ψ_2, the overlapping lobes of the original partners are of opposite signs, and its sign changes as one goes across the interaction region from the center of ϕ_1 to the center of ϕ_2. The surface where the sign of ψ_2 changes is called a nodal surface (plane).

In summary, the bonding combination ψ_1 lacks a nodal plane and has an increased electron density between the atoms (in-phase mixing if $S_{12} > 0$), while the antibonding combination ψ_2 has a nodal plane and a reduced electron density between the atoms (out-of-phase mixing if $S_{12} < 0$). The bonding orbital ψ_1 will accommodate two electrons and thus produce a net stabilization (a chemical bond). This is the origin of the classical "shared electron pair." For maximum bonding stabilization, it is desirable to maximize the absolute value of the overlap integral S_{12}.

A third electron would have to enter the antibonding orbital ϕ_2, causing a loss of much and possibly all of the overall stabilization. A fourth electron would also have to enter ϕ_2, and now interaction would definitely lead to a net destabilization. For this reason, filled AOs (lone pairs) repel and avoid one another ("lone-pair repulsion"). For minimum destabilization, it is desirable to minimize $|S_{12}|$. For instance, if the lone pairs are of the p type and are on adjacent atoms, a dihedral angle of close to 90° will be preferred (e.g., in H_2O_2).

If the original orbitals ϕ_1 and ϕ_2 are degenerate, $\Delta\varepsilon = 0$, they will enter the new orbitals ψ_1 and ψ_2 with equal weights. If the initial energies are unequal, the more stable initial orbital ϕ_1 will enter with a numerically larger coefficient into the more stable resulting orbital ψ_1 and with a smaller coefficient into the less stable resulting orbital ψ_2. The opposite will be true for the less stable initial orbital ϕ_2. This result is quite logical: In the more stable of the two orbitals, the electrons spend more time on the more electronegative atom.

D. Construction of Molecular Orbitals

1. "Localized" Bonds

In the first step, AOs on the participating atoms are mixed to produce hybrids pointing in the directions of the desired bonds. These are then combined pairwise, each pair producing a bonding and an antibonding orbital strictly localized at the atoms to be bound. Occupancy of the bonding orbital by two electrons then produces a localized bond, while the antibonding orbital remains unused. If electrons are left over after all bonding orbitals have been occupied twice, they are placed into remaining AOs or hybrid orbitals that have not been used in the above-mentioned procedure (nonbonding lone pairs or unpaired electrons). So far, the picture resembles the classical description, without resonance.

At a more advanced level, it is recognized that the several bonding and antibonding orbitals located at a single atom have nonvanishing interaction elements with one another. If they are allowed to mix, they will produce the fully delocalized canonical MOs. This usually has a predictable effect on the total energy, which can be absorbed in bond additivity schemes. Exceptions are the case of cyclic delocalization (e.g., in cyclopropane) and certain cases of stabilization of radicals, carbenes, and biradicals. A well-known exception in which such delocalization must be considered involves the interaction of a lone pair on an atom such as oxygen with the orbitals corresponding to bonds leading to its neighbors, which dictates certain stereochemical preferences known as the anomeric effect.

2. Delocalized Bonds

Although delocalized bonds could be obtained in the same manner in principle, it is more efficient to proceed more directly. All n AOs (and possibly hybrid orbitals) that participate in the "delocalized" system are mixed at once, producing n canonical delocalized MOs directly. Typically, these are sets of orbitals characterized by mutual π overlap and referred to as the π-electron system. A good example are the six p orbitals of the sp^2 hybridized carbons of benzene that are left over after the localized bond framework has been constructed (Fig. 4).

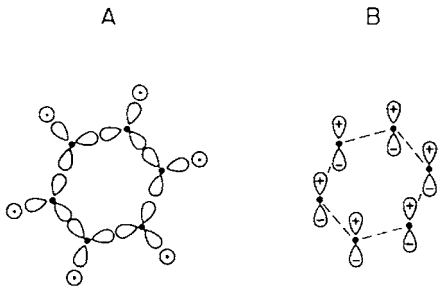


FIGURE 4 (A) Hydrogen Is atomic orbitals and carbon sp^2 hybrid orbitals (major lobes only) needed for the description of "localized" σ bonds in benzene. (B) Carbon $2p_z$ atomic orbitals needed for the description of "delocalized" π bonds in benzene.

On the one hand, the process of simultaneous mixing of more than two orbitals is more difficult to visualize, but on the other hand, there is usually only one participating orbital on an atom, simplifying matters somewhat. Usually, about half of the resulting MOs are stabilized (bonding), perhaps one or a few are nonbonding, and about a half are destabilized and antibonding. Their approximate forms can be guessed from the requirement that a nodal surface is added each time that one goes to the energetically next higher MO. Mnemonic devices for the resulting energy patterns are shown in Fig. 5 for linear and cyclic polyenes. In the latter case, the origin of the special stability ("aromaticity") of $4N + 2$ electron systems is apparent: This is the electron count needed to produce a closed shell. In general, however, at least a back-of-the-envelope numerical

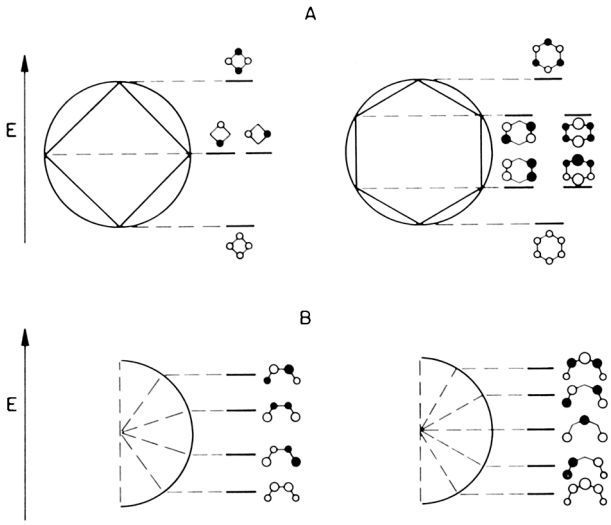


FIGURE 5 Conjugated π systems; orbital energies and shapes (top view, white represents positive, black represents negative sign). (A) Cyclic conjugation (annulenes). (B) Linear conjugation (polyenes).

calculation is necessary. Here, Dewar's perturbational MO (PMO) method is particularly useful.

The interaction of fully or partly localized orbitals is frequently referred to as conjugation. Interactions between several AOs that belong to a π system can be referred to as π conjugation, or conjugation for short. Interactions of an orbital of the π type with one or more suitably aligned bond orbitals of the σ type are referred to as hyperconjugation. Interactions between adjacent bond orbitals of the σ type or one such orbital and a lone-pair orbital are referred to as σ conjugation. As noted above, they occur through the interaction between two hybrid orbitals located on the same atom.

3. Hypervalent Bonding

The procedures described so far will not work for hypervalent atoms that do not have enough valence AOs to provide a sufficient number of localized bond orbitals. Without going into detail, we note that in this case three-center orbitals can be constructed instead, and a satisfactory simple description results.

The expansion of the electron count in the valence shell of an atom of a main-group element beyond eight is a reflection of the inadequacy of the way in which the classical rules of valence assign electrons into a valence shell. In reality, the true total electron density in the region of space corresponding to an atomic valence shell does not reach these high formally assigned numbers. Fundamental limitations on the latter are given by the Pauli principle, which demands that any electron density in excess of an octet has to occupy orbitals of the next higher principal quantum number (Rydberg AOs). It is a gross oversimplification, however, to pretend that the two electrons of a bond contribute full occupancy of two to the valence shell of each of the atoms connected by the bond, particularly when the two atoms differ greatly in electronegativity. The degree to which the valence shell of the atom of the more electropositive element is actually filled is overestimated by the simple rules, and this accounts for the ease with which it enters into hypervalency.

III. QUANTITATIVE ASPECTS OF MOLECULAR STRUCTURE

A. Potential Energy Surfaces

1. Construction of Potential Energy Surfaces

Before attempting a quantitative quantum mechanical treatment of molecules it is customary to separate the motions of nuclei from those of electrons. This separation is known as the Born–Oppenheimer approximation, and it underlies all of the current thinking about molecular structure. It is justified by the much larger mass of nuclei compared with that of electrons, which causes the nuclei to move much more slowly than electrons. Thus, electrons adjust their motions essentially instantaneously to any change in the location of the nuclei, as if they had no inertia.

From the viewpoint of quantum theory a molecule is a quantum mechanical system composed of atomic nuclei and electrons. It has an infinite number of stationary states, that is, states whose measurable properties do not change with time. Each of these states is characterized by an energy and a wave function. A wave function is a prescription for assigning a numerical value ("amplitude") to every possible choice of coordinates for all particles in the system. A square of the number assigned to any choice of these coordinates represents the probability density that a measurement will find the system at that particular collection of coordinates.

In order to find the stationary states of a quantum mechanical system, their energies, and wave functions, one must solve the Schrödinger equation $\hat{H}\psi = E\psi$, where $\hat{H}$ is the Hamiltonian operator of the system, ψ the wave function, and E the energy of the stationary state.

In order to obtain a quantum mechanical description of a molecule within the Born–Oppenheimer approximation, at least in principle, one proceeds as follows. Fixed molecular geometry is assumed. Mathematically, this corresponds to choosing a point in the nuclear configuration space. As soon as this is done, the Hamiltonian operator for electronic motion is fully defined so that the corresponding Schrödinger equation can be solved for ψ and E. An infinite number of solutions exist, differing in their energies, E, and wave functions, ψ.

One of the characteristics of an electronic wave function is the number of unpaired electrons it contains. Those wave functions in which this number is zero describe singlet states. Those that contain two unpaired electrons describe triplet states, and so on. Wave functions of radicals can have one unpaired electron (a doublet state), three (a quartet state), and so on.

Of the infinite number of stationary wave functions that are solutions to the Schrödinger equation for a chosen nuclear geometry, one is of lowest energy. Almost always this is a singlet wave function if the organic molecule has an even number of electrons and a doublet wave function if it has an odd number of electrons. In the following, we assume an even number of electrons. We label the lowest energy singlet wave function $\psi(S_0)$ and its energy $E(S_0)$. The next higher energy singlet wave function is then identified and labeled $\psi(S_1)$, and its energy is identified and labeled $E(S_1)$. This is the wave function of the first excited singlet state. Similarly, the wave functions of the second

and higher electronic excited singlet states can be identified. Among the triplet wave functions the one with the lowest energy is called $\psi(T_1)$ and its energy $E(T_1)$. Similarly, higher triplet state wave functions and their energies are identified.

A different molecular geometry is then chosen and the process repeated. While this is difficult to do in practice, one can at least imagine performing this kind of operation for all possible molecular geometries. In a plot of $E(S_0)$ against the values of the geometrical parameters that describe the molecular structure, a surface will then result. This is the potential energy surface for this particular electronic state of the molecule (the potential energy of the molecule is its total energy minus the energy of overall translational, rotational, and vibrational motion).

The resulting surface is easy to visualize if only one or two geometrical variables are used to describe the molecular structure. As shown in Fig. 6, in the former case the set of points $E(S_0)$ represents a line; in the latter case it represents a two-dimensional surface, often displayed in the form of a contour diagram. For all organic molecules of real interest, the number of independent geometrical variables necessary for the description of the internal geometry is large ($3N-6$, where N is the number of atoms). The resulting surfaces are multidimensional and difficult to envisage. Frequently, they are referred to as hypersurfaces.

What can still be visualized readily are one-dimensional or two-dimensional cross sections through these hypersurfaces, which correspond to only a limited variation of molecular geometries, particularly to specific kinds of intramolecular motion, such as rotation around a bond.

2. Motions on the Surfaces

The gradient of the potential energy surface defines the forces acting on the nuclei. The resulting changes of molecular shape can be represented by a point that moves through the nuclear configuration space. It is possible to visualize the vibrational motions of the molecule as well as its internal rotations as the motions of a marble rolling on the potential energy surface.

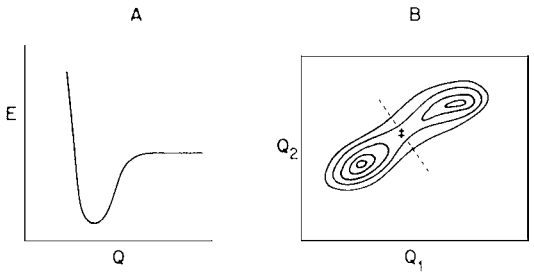


FIGURE 6 Display of a one-dimensional (A) and a two-dimensional (B) cut through a potential energy surface. In (A), energy is plotted against a geometrical variable Q. One minimum is present. In (B), contour lines connect points of equal energy in the Q_1, Q_2 geometrical space. Two minima are present. The separation into two catchment basins is shown by a dashed line; the transition state structure is indicated by a double dagger.

If the molecule is isolated, its total energy will remain constant and the marble will perform endless frictionless motion on the surface, trading the potential energy against the kinetic energy of nuclear motion and vice versa. If the molecule exchanges energy with its environment, it will tend to lose any excess energy it may have and settle in one of the valleys or minima on the surface.

A proper description of the motion that corresponds to vibrations and internal rotations again must be quantum mechanical since even the relatively heavy nuclei really obey quantum rather than classical mechanics. Once again, one can find the stationary states of the vibrational motions and their wave functions and energies by setting up the appropriate Schrödinger equation and solving it. The Hamiltonian operator that enters into this equation now contains the information on the potential energy embodied in the shape of the potential energy surface. An infinite set of possible solutions again exists. A finite number of solutions have energies corresponding to bound states, that is, those with energies below the dissociation limit for the molecule (energy required to break the weakest bond and separate one of the atoms to infinity). The wave function of lowest energy represents the vibrational ground state of the molecule. In this state the kinetic energy of the nuclear motion is not zero since this would violate the uncertainty principle. It is referred to as the zero-point energy.

Most molecules have a well-defined equilibrium geometry that corresponds to a minimum in the $E(S_0)$ surface. The wave function of the lowest vibrational stationary state is heavily localized near this minimum. If the shape of the potential energy hypersurface in this vicinity can be approximated by a paraboloid, the vibrational motion in the lower vibrational states is harmonic and can be described as a product of $3N-6$ normal mode motions ($3N-5$ for a linear molecule), each characterized by a frequency ν. The zero-point energy is obtained by summing the contribution $\frac{1}{2}h\nu$ from each of the normal modes. The value $h\nu$ is equal to the energy separation from the lowest to the next higher energy level in each mode. The transitions between these individual levels lie in the region from several hundred to several thousand wave numbers and are commonly studied by infrared and Raman spectroscopy.

In most cases more than one local minimum is found on the $E(S_0)$ hypersurface (Fig. 6B). This means that a given collection of nuclei and electrons has more than one possible equilibrium geometry. Usually, this means that several isomers of the molecule exist, but some of the minima may

also correspond to dissociation products, the sets of two or more smaller molecules formed from the same collection of nuclei and electrons. Some of the minima may have equal energies (e.g., a pair of enantiomers). In general, one minimum is of lower energy than all the others and corresponds to the most stable isomer of the molecule.

Each minimum is surrounded by its catchment basin (Fig. 6B). Within a given basin, the potential energy surface slopes toward the minimum in question. Each catchment basin defines the range of geometries that correspond to a given chemical species.

3. Chemical Reactions

Under ordinary conditions molecules are frequently not in stationary vibrational states. Vibrations and internal rotations are affected by collisions with neighboring molecules, which add or subtract small amounts of vibrational energy more or less randomly. At thermal equilibrium the vibrational energy content is $\frac{1}{2}kT$ for each degree of freedom. It is this supply of random thermal energy that permits molecules to escape from one catchment basin to another in a thermal reaction.

In order to move from one minimum to another it is necessary to overcome ridges that separate them (Fig. 6B). This is done most easily by travel through saddle points, which correspond to transition structures and are usually called transition states. Some of these lie only a little higher in energy than a starting minimum. Travel over these saddles is easy even at low temperatures, and a chemical species corresponding to a shallow catchment basin may not be isolable at room temperatures or even below. Conformational isomers are a good example of such chemical species. Other catchment basins may be deeper and the surrounding saddle points difficult to reach. They often correspond to configurational or constitutional isomers.

An important attribute of a saddle point connecting two catchment basins is the width of the saddle. Very narrow saddles are difficult for the rolling marble to find and decrease the probability of travel from one catchment basin to another. Wide saddles accommodate a much larger flux under otherwise identical conditions and lead to fast motion.

For a reaction whose rate is not limited by the rate of diffusion, the reaction rate is well described by Eyring's transition state theory. For the rate constant k, we have the following:

$$k = \kappa \frac{KT}{h} \exp\left(\frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT}\right),$$

where κ is the transmission coefficient, k the Boltzmann constant, T absolute temperature, h Planck's constant, R the gas constant, $\Delta H^{\ddagger}$ the activation enthalpy (i.e., the difference from the lowest vibrational level in the original catchment basin to the lowest level available over the saddle point), and the activation entropy $\Delta S^{\ddagger}$ is a function of the width of the saddle relative to that of the starting minimum.

Isotopic substitution does not have any effect on the potential energy surfaces as usually defined. However, it does affect the dynamics of the vibrational process and the dynamics of the motion over saddle points by changing the magnitude of the energy of zero-point motion and the spacing of the vibrational stationary states and thus the density with which these states are packed. Isotopic substitution thus leads to changes in vibrational spectra and also to changes in reaction rate constants.

Some vibrational stationary states have energies that lie above some of the lower saddle points. In such states the molecule is free to travel from one catchment basin to another and in that sense has no fixed chemical structure. This sort of situation occurs, for instance, for rotations around single bonds at elevated temperatures.

The perfect quantum mechanical analogy to the rolling-marble description given earlier is described in terms of the motion of a wave packet, that is, of a nonstationary wave function initially more or less strongly localized in a particular region of nuclear geometries. Since the vibrational levels are spaced quite closely together, however, the classical description in terms of the rolling marble is often quite adequate.

In the Born–Oppenheimer approximation, the overall rotation of a molecule can also be uncoupled from other kinds of motion. For an isolated molecule, the solution of the Schrödinger equation for rotational motion leads to a set of very closely spaced stationary levels. Transitions between these levels occur in the microwave region of the electromagnetic spectrum. In solution, the quantization of rotational motion is usually destroyed by intermolecular interactions, so that it has little importance in theoretical organic chemistry.

So far we have concentrated on the lowest singlet hypersurface $E(S_0)$. However, motion can be studied similarly on other hypersurfaces if they can be calculated to start with. Such electronically excited states are usually produced either by absorption of light in the ultraviolet and visible regions, as studied by electronic spectroscopy, or by energy transfer from another electronically excited molecule. The latter is particularly useful in the case of excitation into a triplet state, since excitation from a ground singlet to a triplet state by direct absorption of light has a very low probability.

The study of the processes involving the higher potential energy surfaces is the domain of photochemistry and photophysics. These are very difficult or impossible to understand in terms of the classical picture described in

Section I, and it is thus easy to see why photochemists are more likely than any other organic chemists to discuss chemistry in terms of potential energy surfaces. Further discussion of photochemically induced reactions can be found below.

4. Reaction Coordinate Diagrams

For many purposes it is useful to condense the multidimensional complexity of molecular motion from one catchment basin to the next into a one-dimensional reaction coordinate diagram in which the degree of progress from the first minimum over the transition state to the second minimum is plotted horizontally as the so-called reaction coordinate. The quantity plotted vertically can be the potential energy E, which we have been discussing so far, but then all information that has to do with the properties of the surface along dimensions other than the reaction coordinate is lost. It is more common to plot instead the Gibbs free energy $\Delta G = \Delta H - T\Delta S$ corresponding to all degrees of freedom other than the reaction coordinate (ΔH is enthalpy, ΔS is entropy). In this manner, information on the entropic constraints dictated by the shape of the potential energy surface in directions perpendicular to the reaction path is preserved.

The effects of structural perturbations on such reaction diagrams can be relatively easily envisaged. For instance, in many cases a structural factor that will stabilize the product relative to the starting material will also tend to stabilize the transition state, albeit to a smaller degree, as is indicated schematically in Fig. 7. Reactions of this kind are said to follow the Bell–Evans–Polanyi principle. Such reactions tend to obey linear free energy relationships, which tie thermodynamic quantities such as equilibrium constants to rate quantities such as kinetic rate constants: $\Delta G_1^{\ddagger} - \Delta G_2^{\ddagger} = \alpha(\Delta G_1^0 - \Delta G_2^0)$, where $\Delta G^{\ddagger}$ is the free energy of activation and ΔG^0 the free energy of reaction. Perhaps the best known example is the Brønsted law, which relates equilibrium and kinetic acidity. In general, linear free energy relationships interrelate changes in free energies or free energies of activation for a series of reactants, usually differing by substitution. They are useful in mechanistic studies.

Cases are also known in which structural perturbations act quite differently on the transition state and the product. Then, reactions with larger equilibrium constants do not necessarily proceed faster than those with smaller ones.

Another consequence of a parallel stabilization of a product and of the corresponding transition state is the displacement of the transition structure toward the starting materials along the reaction coordinate (Fig. 7). Those reactions that obey the Bell–Evans–Polanyi principle thus will have earlier transition states relative to other, similar reactions if they are more favorable thermodynamically and a later transition state if they are less favorable. This statement is known as the Hammond postulate.

At times it is useful to separate motion not only along the reaction coordinate, but also along one other direction selected from all the others and to plot the free energy of the reacting system against two geometrical variables. Diagrams of this kind are particularly popular in the study of substitution and elimination reactions and are known as More O'Ferrall diagrams. The geometries corresponding to the starting material and the product lie at diagonally opposed corners of a square (Fig. 8) and are connected by an energy surface that rises up to a saddle point and then descends to the other corner. The effects of a change in the relative stability of the starting materials and products

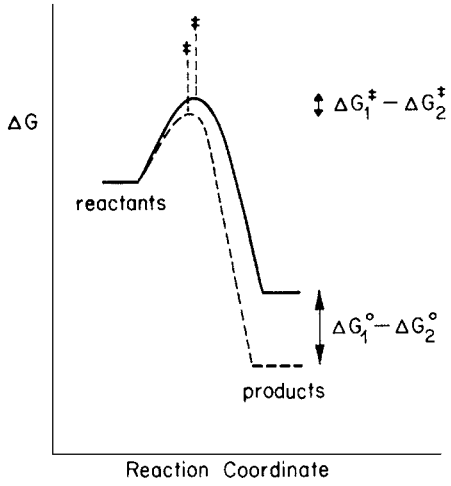


FIGURE 7 Reaction coordinate diagram for two similar reactions differing in ΔG. Location of the transition states is indicated by double daggers.

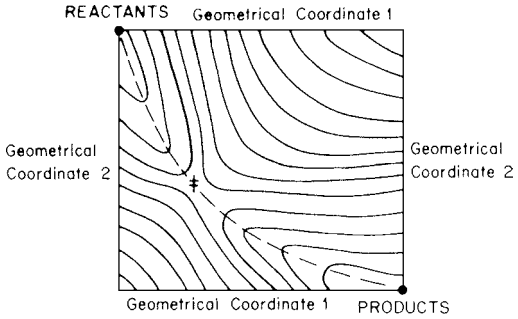


FIGURE 8 More O'Ferrall contour diagram for a reaction requiring two types of nuclear motion that can proceed more or less synchronously. The reaction coordinate is shown as a dashed line, and the structure of the transition state is indicated by a double dagger. The two motions involved could be, for instance, the approach of a nucleophile and the departure of the leaving group in a nucleophilic substitution reaction.

on the position of the transition state are given by the Hammond postulate: The transition structure moves away from the corner that has been stabilized. The shape of the surface along the other diagonal is just the opposite: The reaction path is of low energy relative to the two corners. For this reason a change in the relative stability of the geometries represented by the two corners has an effect opposite to that discussed in the Hammond postulate. A stabilization of one of the corners will move the transition point closer to that corner.

While it can be generally assumed that the molecules in the initial catchment basin are almost exactly at thermal equilibrium, which is being perturbed quite insignificantly by the escape of the most energetic molecules over a saddle point or saddle points, this is not always true. In hot ground state reactions a molecular assembly is initially generated with a vibrational energy content much higher than would correspond to the temperature of the surrounding medium, perhaps as a result of being born in a photochemical process or in a very highly exothermic thermal process. Such an assembly can then react at much higher rates.

A study of the reaction rates of molecules starting in individual vibronic levels provides much more detailed information than the study of molecules that are initially at thermal equilibrium but has so far remained a domain of small-molecule chemical physicists and has had little impact on the theory of organic reactions.

An important point to consider is the possibility that a molecule has a choice of more than one saddle over which it can escape from an original catchment basin. The nature of products isolated from such competing reactions can depend on the choice of reaction conditions. Under kinetic control the molecules of the two products are not given an opportunity to travel back over their respective saddles into the initial catchment basin. The preferred product will be that which is being formed faster. On the other hand, if the molecules are provided with an opportunity to return to the initial catchment basin so that an equilibrium is eventually approached, the reaction is said to be run under thermodynamic control and the thermodynamically more stable product will be isolated.

B. Electronic Wave Functions

Once a wave function of a quantum mechanical state is available, it is possible to calculate its observable properties in a straightforward manner. Unfortunately, the Schrödinger equation is very difficult to solve, and this has not been achieved exactly for any but the simplest molecular systems. The fundamental mathematical problem can be traced to the fact that even for fixed nuclear positions the molecule still represents a many-body system if it contains more than one electron. Nevertheless, it is possible to obtain useful approximate solutions. These become gradually less accurate as the molecule in question increases in size.

Essentially all of the approximate methods for the calculation of electronic wave functions for an organic molecule start with a basis set of AOs mentioned above in the qualitative discussion. In a so-called minimum basis set description, the AOs assigned to an atom in a molecule are those that would be occupied in a neutral isolated atom of that element plus the remainder of its valence shell. The AOs can be used in two ways for the construction of molecular electronic wave functions.

1. The Valence–Bond Method

In the conceptually simple but computationally very difficult valence–bond (VB) method one first assigns all available electrons to individual AOs in a way that satisfies the Pauli principle; that is, placing no more than two electrons in any one AO and assigning opposite spins to electrons that share the same AO. The method of doing this is not described in detail here, but it is possible to ensure that the resulting many-electron wave function is properly spin-adapted and represents a singlet, a doublet, a triplet, and so on, as appropriate. An example is the H^+H^- VB structure for the H_2 molecule, in which the two available electrons are both assigned to the 1*s* orbital of the hydrogen on the right. Since they are then necessarily spin-paired this represents a singlet structure.

It is generally possible to write a very large number of such VB structures for any given molecule. For instance, for the H_2 molecule we could equally well have written H^-H^+ or four additional structures in which one electron was assigned to each of the available 1*s* AOs and which differed only in the assignment of electron spin. In order to obtain the desired final electronic wave function ψ it is necessary to allow these VB structures to mix.

Mathematically, this corresponds to taking a linear combination of all structures of the same spin multiplicity. In order to determine the coefficients with which each of the structures enters into the final mixture, one uses the variational theorem, which states that the best approximation to the ground wave function $\psi(S_0)$ is given by those coefficients that lead to a wave function of lowest energy. Determining the coefficients requires the diagonalization of a matrix the dimension of which is equal to the number of VB structures of a given multiplicity possible for the molecule. This increases very rapidly with the size of the molecule, making a full calculation impractical even for quite small organic molecules. Another important difficulty in the calculation is the complicated nature of the computation of each of the elements of the matrix, due ultimately to the fact that the AOs are not orthogonal.Because

of these mathematical difficulties the VB method has not had much numerical use, although it remains important as a conceptual tool.

2. The Molecular Orbital Method

The way in which almost all computations of molecular electronic wave functions are performed nowadays is the MO method already mentioned. In this method it is recognized from the outset that in a molecule electrons are delocalized over the whole region of space spanned by the nuclear framework. Accordingly, AOs are first combined into MOs. Unlike the VB structures mentioned above, these are still one-electron wave functions.

Mathematically, the MOs are written as linear combinations of atomic orbitals. From n AOs it is possible to construct up to n linearly independent MOs. This is more than is usually needed since in the next step, in which electron occupancies are assigned to these molecular orbitals, each can be occupied by up to two electrons of opposite spins. The unused MOs are referred to as unoccupied (or virtual). Such an assignment of electron occupancies to MOs results in a many-electron wave function known as a configuration. The number of configurations possible is usually very large and is in fact equal to the number of VB structures possible for the same spin multiplicity (singlet, doublet, etc.). In order to obtain the final electronic wave function ψ it is now necessary to mix all possible MO configurations in a fashion analogous to that used to mix the VB structures before. This configuration mixing is referred to as configuration interaction (CI). Indeed, the final wave function ψ obtained by the VB procedure in which all VB structures are included and that obtained by the MO procedure in which all configurations are included will be one and the same. It is referred to as the full configuration interaction (FCI) wave function. It cannot be obtained in practice for any but the smallest organic molecules because the number of possible configurations increases astronomically with molecular size.

The MO path to the FCI wave function ψ involves one more step than the VB path, and this appears to be more complicated at first sight. However, the availability of the additional step endows the MO method with more flexibility than the VB method, which permits a mathematically much simpler formulation. This flexibility arises in the step of combining AOs in MOs. First, it is easy to ensure that the MOs are mutually orthogonal, and this leads to an immense simplification in the calculation of matrix elements. Second, it is also possible to optimize the choice of the MOs in such a way as to speed up the convergence of the final summation in which configurations are combined to obtain the state wave function.

Indeed, in the most commonly used form of the MO procedure this final step is omitted altogether, and a single configuration built from optimized molecular orbitals is used as an acceptable, if poor, approximation of the FCI electronic wave function ψ. These optimum MOs are known as the self-consistent field (SCF) or Hartree–Fock (HF) MOs. They are obtained using the variational principle and demanding that the energy of the one configuration under consideration be as low as possible. The energy difference between the HF description and the FCI description is referred to as the correlation energy. It is important to note that weak intermolecular interactions (van der Waals interactions), important in processes such as molecular recognition and complexation, cannot be calculated at the SCF level. In calculations of these effects, inclusion of correlation effects is essential.

One of the advantages of the MO method is that it is relatively easy to improve the SCF solution partially without having to go all the way to the unreachable limit of FCI. Several methods for making such an improvement are available, such as (1) limited configuration interaction in which the CI expansion is truncated in some systematic fashion well before the FCI limit is reached and (2) methods in which correlation effects are viewed as a small perturbation of the SCF solution and are treated by perturbation theory. The use of these methods is particularly important (1) when the electronic wave function ψ is being calculated for a geometry far removed from the molecular equilibrium geometry, (2) when the molecule has very low lying excited electronic states, (3) when the molecule is a biradical, (4) if the calculation is performed for an excited electronic state, or (5) if intermolecular forces are to be calculated.

So far we have discussed only the minimum basis set approximation. This might be quite adequate if the AOs were chosen in a truly optimal manner, but it would require a nonlinear optimization and this is normally not done. Even so, in order to obtain a truly accurate solution for the electronic wave function ψ within the Born–Oppenheimer approximation it would be necessary to increase the number of basis set AOs used in the calculation (in principle, to infinity).

The form of the AOs usually adopted in numerical work is normally chosen for computational convenience (Gaussian-type orbitals) in a way that makes them quite different from the optimum orbitals of an isolated atom. Thus, in practice most computations are not performed with a minimum basis set but with extended basis sets, and several such standard sets are in common use.

The electronic wave functions resulting from such large-scale calculations are usually not easy to visualize. Frequently, it helps to draw the resulting electron densities in the form of contour maps. Alternatively, electrostatic

potential contour maps can be constructed, indicating the nature of the electric fields to be expected in the vicinity of the molecule.

3. *Ab Initio* and Semiempirical Methods

The approach described so far is of the so-called *ab initio* type, in which the whole computation is done from first principles, taking from experiment only the values of fundamental constants such as electron charge. As already indicated, accurate results cannot be obtained by these methods for most molecules of interest in organic chemistry, but the approximate solutions obtained at the SCF or improved SCF level provide much useful information. For certain properties, such as molecular geometries, dipole moments, and the relative energies of conformers, the agreement with experiment is excellent.

An alternative approach to the problem of molecular electronic structure is provided by semiempirical models. In these no attempt is made to derive the properties of atoms from first principles. Rather, they are taken as described by a set of parameters obtained by fitting experimental data, and an attempt is made to find a model Hamiltonian that will provide a good description of interatomic interactions. The form of the model Hamiltonian is patterned after the *ab initio* analysis. In the most common semiempirical methods it is still a fairly complicated many-electron Hamiltonian, so that its exact stationary wave functions cannot be found for molecules of interest, and only approximate solutions are obtained. One almost invariably starts with a minimum basis set of AOs and proceeds to an SCF type of wave function, possibly followed by a limited amount of improvement toward the FCI wave function. The parameters that enter the Hamiltonian are optimized so as to bring about close agreement between the molecular properties computed from the approximate wave function (usually SCF) and those observed experimentally. Most commonly, the properties fitted are heats of formation, molecular equilibrium geometries, or suitable spectroscopic properties. In this way one attempts to incorporate intraatomic correlation energies and a large part of interatomic electron correlation energies into the model through parameter choice, although one works only at the SCF level or at least not much beyond it.

The agreement of the calculated properties with experiment is roughly comparable to the agreement obtained by extended basis set *ab initio* methods at the level of the SCF approximation or slightly better, at least for those classes of molecules for which the semiempirical parameters were originally optimized. However, even for molecules quite different from those on which the original optimization was performed, the agreement is frequently striking considering that orders of magnitude less computer time is needed for the semiempirical computations. This great reduction in computational effort is to a large degree due to the almost universal use of the so-called zero differential overlap approximation, which greatly reduces the number of electron repulsion integrals needed in the computations.

Some of the best known examples of semiempirical methods are those developed for the treatment of electronic ground states by Dewar and collaborators (MNDO, AM1, and MINDO/3) and the methods developed by Jaffé (INDO/S), Zerner (INDO/S), and their respective collaborators for calculations involving electronically excited states. A very simple procedure is the extended Hückel method popularized by Hoffmann. Others are the older methods developed for the treatment of π electrons only: the PPP method of Pariser, Parr, and Pople and the extremely crude but also extremely simple HMO method of Hückel.

These semiempirical models should not be confused with approximate models that are designed to mimic the results of *ab initio* calculations in a simpler manner rather than to mimic the results of experiments. The best known approximate MO methods are the CNDO and INDO methods developed by Pople and collaborators.

C. Molecular Properties

Although we have indicated how the electronic wave function of an electronic state and its energy are calculated, we have said very little about the calculation of other molecular properties once the wave function is known.

The calculation of molecular equilibrium geometry in a given electronic state, usually S_0, is performed by varying the assumed nuclear geometry and repeating the calculation of the energy by one of the methods referred to earlier until a minimum is found. This search is normally performed by computer routines that compute surface gradients in order to speed up convergence toward a local minimum in the $(3N - 6)$-dimensional nuclear configuration space. From the computed curvatures of the surface at the minimum one obtains the force constants for molecular vibrations and the form of the normal modes of vibration. For a true local minimum, all the force constants must be positive. A similar type of procedure, minimization of the norm of the gradient, can be used for finding transition states. After a transition point is found, it is essential to convince oneself that a normal mode analysis produces only one vibration with a negative force constant. This mode corresponds to the path from one catchment basin to the other, and the corresponding vibration has an imaginary frequency (the restoring force is negative). Other modes of vibration are ordinary and permit the evaluation of the entropy of the transition state.

Most other molecular properties are normally evaluated only at the equilibrium geometry, although strictly speaking they should be calculated at a large number of geometries and averaged over the vibrational wave function of the state in question. Generally, they are obtained by representing the observable by a quantum mechanical operator and computing the expectation value of this operator over the wave function.

At the HF level, *ab initio* or semiempirical, some of these properties can be obtained in an approximate manner more simply. Thus, the lowest ionization potential of the molecule is approximately equal to the negative of the energy of the highest occupied molecular orbital (HOMO). The electron affinity of the molecule is similarly approximated by the negative of the energy of the lowest unoccupied molecular orbital (LUMO). The energies of electronic excitation are usually more complicated to obtain in that the introduction of CI may be quite necessary.

Those electron excitations that can be well described as a promotion of an electron from one single occupied MO to one unoccupied MO are rare but exist in some molecules. An example is the so-called L_a band in the absorption spectra of aromatic hydrocarbons and the first intense band of polyenes. In the SCF approximation the electronic excitation energy is equal to the energy difference between the orbital out of which the promotion occurs and the orbital into which it occurs, minus the repulsion energy of two electron densities, each provided by an electron in one of these two MOs. This approximates the energy of the triplet excited state. The energy of the singlet excited state is higher by twice the exchange integral between the two MOs involved (the self-repulsion energy of a charge density produced by taking a product of the two orbitals).

Due to the approximate nature of the SCF treatment and to the additional approximations involved in the statements just made, the results are usually more useful for an interpretation of trends within a group of compounds rather than the absolute value for any one compound. The semiempirical methods in particular have had much use in this kind of application.

A property related to this is the formation of intermolecular complexes characterized by a charge-transfer transition, which frequently occurs in the visible region and in which an electron is transferred from one molecule to another. The energy of this transition is related to the ionization potential of the donor and the electron affinity of the acceptor moiety and once again can be correlated with the computed MO energies.

Not only the energies but also the coefficients of the MOs computed in the SCF picture are approximately related to observable properties. Thus, for a given MO, the square of the coefficient on a particular AO is related to the probability that an electron in that MO will be found in that particular AO. This relation is particularly simple in those semiempirical methods that use the zero differential overlap approximation. Then, the AOs are mutually orthogonal and the relation between the square of the coefficient and the probability is a simple proportionality. In methods that use nonorthogonal AOs, such as the *ab initio* ones, it is more difficult to define electron populations for AOs and atoms. The procedure usually used is known as the Mulliken population analysis. This permits a calculation of electron densities in AOs and of total electron densities on atoms. These in turn can be related to molecular dipole moments, infrared spectral intensities, and, much more approximately, to nuclear magnetic resonance (NMR) shielding constants.

The distributions of unpaired spin obtained in an analogous fashion from the squares of coefficients of a singly occupied orbital are related to the hyperfine coupling constants in electron spin resonance spectroscopy.

There is another class of molecular properties that are not related in a simple way to the expectation value of an operator: the so-called second-order properties. Some of the most important of these are molecular polarizability, Raman intensities, and chemical shielding in NMR spectroscopy. They can be computed by introducing an outside perturbation such as an electric or magnetic field explicitly into the calculation of the molecular wave function. These calculations are more difficult and less reliable, particularly with respect to magnetic properties, where the incompleteness of the basis sets used tends to make the results dependent on the choice of origin of coordinates. Good progress has been made in the calculation of NMR chemical shielding constants and their anisotropies, while the calculation of intensities in Raman spectra still leaves much to be desired.

IV. REACTION PATHS

A. Thermal Reactions

A reaction involving motion through a single transition state is referred to as an elementary reaction step. Most reactions of organic compounds involve a sequence of such reaction steps in which the reacting system passes through a series of intermediates and transition states that separate them. The reaction intermediates are usually unobserved. An important part of the investigation of organic reactivity is the determination of these individual steps in an overall sequence, usually accomplished using the tools of chemical kinetics.

In terms of transition state theory an understanding of the reactivity of organic molecules under conditions

usually employed in the laboratory requires the knowledge and understanding of activation enthalpies and activation entropies for elementary reaction steps; that is, the shapes of the relevant potential energy surfaces in the vicinity of the local minimum in the reactant catchment basin and in the vicinity of the transition state saddle point in the ridge separating it from the product catchment basin.

The reaction path of an elementary step is usually defined as the steepest descent path from the transition point to the minima in the starting and the final catchment basins. It describes the geometry of the approach of the reacting molecules and, specifically, of the reaction centers toward one another. This approach is frequently described in pictorial terms, such as "face-to-face" approach of an olefin to a diene in a Diels–Alder reaction or a "backside attack" by a nucleophile on a carbon atom carrying a leaving group. Frequently, the geometrical path is described in even more quantitative terms—for example, by giving an angle of approach of a nucleophile attacking a carbonyl group.

Note, however, that within the constraints of the transition state theory only the geometry at the transition state and in its immediate vicinity matters, and the preceding path taken by the molecules to approach this point is irrevelant. Molecules do not react by following a well-described straight path of minimum energy toward the transition state. Rather, they undergo random excursions in their shape, mutual orientation, and distance that have little to do with any particular path and are dictated by collisions with the surrounding medium. This chaotic motion continues until the molecules happen to acquire enough energy and the appropriate direction of motion to reach the transition state, and then passage to the product catchment basin (i.e., a chemical reaction) follows.

Although the concept of a reaction path therefore cannot be taken as literally representing the actual movements of a reacting molecule, it is useful in that it helps to understand the geometries at which transition states occur. In most reactions other than mere conformational changes, bonds are broken and made (either directly or by means of electron transfer). If old bonds had to be broken completely before new ones were made, the enthalpies of activation would be given by bond strengths. In reality, they are frequently lower. The search for favorable energy paths and low-energy transition states is thus equivalent to a search for ways in which the transition state can be stabilized by introducing new bonding interactions before the old ones have been completely lost.

Two important ways of introducing favorable interactions between the reacting centers can be visualized from the knowledge of the electronic wave functions of the reactants. One of these is to bring unlike charges on the two partners close together and keep like charges well separated. The other is to introduce favorable interactions of occupied MOs of one reacting center with the unoccupied ones of the other and vice versa. Often a reaction can be viewed as an interaction of a Lewis acid with a Lewis base. Then the former kind of favorable interactions prevails in the reaction of a so-called hard acid with a hard base, while the latter kind prevails in the interactions of a so-called soft acid with a soft base. Interactions of a hard partner with a soft one are generally less favorable.

In the frontier MO theory of organic reactivity, introduced by Fukui, attention is limited to the HOMO and the LUMO of the reacting partners. Usually, their energy separations are such that only one of the HOMO–LUMO interactions has to be considered, namely that between the higher lying of the two HOMOs (that of the better donor molecule) and the lower lying of the two LUMOs (that of the better acceptor molecule). A favored orientation for the reaction will be one in which the interaction of these two MOs is maximized.

Very helpful tools for identifying favorable reaction paths and recognizing superficially similar unfavorable ones are correlation diagrams. In their most useful form these resemble crosscuts through potential energy hypersurfaces and indicate how electronic states of starting material correlate with those of a product. Usually, the easiest way to construct such state correlation diagrams is to proceed in several steps starting with an MO correlation diagram, proceeding to a configuration correlation diagram, and finally introducing configuration mixing to obtain the desired state correlation diagram.

The use of molecular symmetry is helpful in the construction of these diagrams because this frequently permits ready identification of the nodal patterns in the wave functions involved, which determine the behavior of the energy along the reaction path. Since nodal patterns are not particularly sensitive to minor perturbations, it is often possible to ignore those secondary perturbations that lower the symmetry of the reacting system and still obtain useful results. For example, the face-to-face interaction of ethylene with propylene can be understood on the basis of considering the interaction of two ethylene molecules instead.

The use of correlation diagrams has been particularly helpful in the case of the so-called pericyclic reactions. These are elementary reaction steps the transition state of which is characterized by a cyclic array of mutually overlapping and interacting AOs on the reacting centers. It is not necessary and usually not possible for all the interactions along the periphery of the cycle to be equal in magnitude.

Such transition states are isoelectronic with cyclic conjugated π systems and can therefore be classified as aromatic or antiaromatic. Not surprisingly, other factors being equal, the aromatic transition states are favored energetically over the antiaromatic ones. The differences are the

largest for small cycles and decrease as the number of AOs in the cycle increases. It has become customary to refer to the pericyclic reaction paths proceeding through aromatic transition states as "allowed" and to those proceeding through antiaromatic transition states as "forbidden." Pericyclic reactions of both types can occur, but the former are normally strongly favored if all other factors are the same. Many different theoretical approaches, of which we mention only two here, have been used to derive simple rules that enable one to predict whether a particular pericyclic process will be of the allowed or forbidden kind. The rules have become known as the Woodward–Hoffmann rules, although many other workers have also made fundamental contributions to their formulation.

Although it is possible to distinguish allowed from forbidden pericyclic reaction paths readily using the aromaticity criterion for the transition state, developed originally by Dewar and Zimmerman and just outlined, correlation diagrams are also often used for this purpose. They are particularly advantageous in the consideration of photochemical processes (see Section IV.B). In order to construct an MO correlation diagram, the MO energies of the reactant are plotted vertically on one side and those of the products on the other side. They are identified as symmetric (S) or antisymmetric (A) with respect to those symmetry elements that are preserved through the whole assumed reaction path and that cut through the bonds being formed or broken. The energies of those MOs that have equal symmetries on both sides are then connected, taking account of the noncrossing rule. This rule states that lines corresponding to wave functions of like symmetries must not cross. The result is shown in Fig. 9 for the face-to-face cycloadditions of two ethylenes and of ethylene with butadiene.

In the latter case (Fig. 9B), all occupied and bonding orbitals of the reactant electronic ground state remain bonding throughout the reaction path and in the product as well. Similarly, all antibonding unoccupied MOs of the starting electronic ground state material remain antibonding and unoccupied. Clearly, bonding is preserved throughout the reaction path, and one would expect the transition state to be of relatively favorable energy. Indeed, the transition state is of the aromatic type, containing six electrons in the cyclic area of orbitals, with positive overlaps of all the orbitals (isoelectronic with the π system of benzene).

On the other hand, in the former case of two ethylenes one of the originally occupied bonding MOs of the reactant becomes antibonding and unoccupied in the ground electronic state of the product, and one of the antibonding and unoccupied orbitals of the reactant becomes bonding and occupied in the ground state of the product. In the region of transition state geometries halfway through the reaction path, both orbitals are approximately nonbonding. Between them, they contain two electrons, and these two electrons do not contribute to bonding in the molecule at the transition state geometry. In effect, the molecule is a biradical and contains one less bond than its number of valence electrons would in principle allow it to have. The transition state is unfavorable and is of the antiaromatic type, containing four electrons in an array of four AOs with all positive overlaps (isoelectronic with cyclobutadiene).

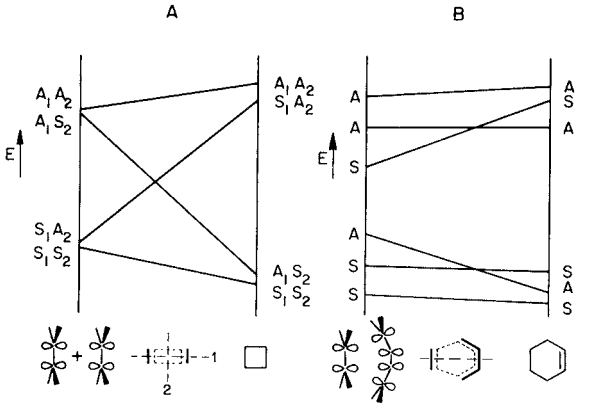


FIGURE 9 Molecular orbital correlation diagrams for the concerted face-to-face cycloaddition of two ethylene molecules (A) and for the concerted Diels–Alder cycloaddition of ethylene with butadiene (B). Symmetry planes preserved throughout the reaction path are indicated by dashed lines at the bottom. Their numbering on the left is keyed to the subscript on the symmetry symbols S (symmetric) and A (antisymmetric) on the molecular orbitals.

Although it is already apparent which of the two reactions chosen as examples is allowed and which is forbidden, it is useful to consider the construction of the configuration correlation diagram as well (Fig. 10). Here

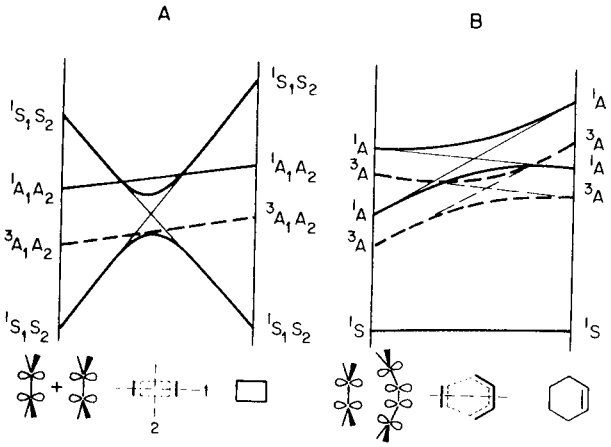


FIGURE 10 Configuration (thin lines) and state (thick lines) correlation diagrams for the concerted face-to-face cycloaddition of two ethylene molecules (A) and for the concerted Diels–Alder cycloaddition of ethylene to butadiene (B). Full lines, singlets; dashed lines, triplets. See legend to Fig. 9.

low-energy configurations are constructed by considering all suitable occupancies of the MOs of the reactants on the left-hand side and of the products on the right-hand side. The symmetry of each is again identified as antisymmetric or symmetric with respect to each of the above symmetry elements, using the rule $S \times S = A \times A = S$ and $A \times S = A$. Correlation lines are now drawn from left to right by keeping the occupancy of each MO in each configuration constant. This often produces crossings of configurations of the same symmetry. According to the noncrossing rule these must ultimately be avoided.

This is accomplished by introducing configuration interaction, which converts the configuration correlation diagram to the desired state correlation diagram, as indicated in Fig. 10. Clearly, the crossing of correlating MOs in the case of the ethylene + ethylene cycloaddition causes a similar crossing of lines in the configuration correlation diagram. Since the effects of configuration mixing, which produces the final state diagram, are generally relatively small, a memory of the crossing at the geometry of the transition state survives and results in a large barrier in the energy of the ground state in the middle of the correlation diagram. It is then concluded that the transition state is unfavorable relative to the case of the ethylene + butadiene process, in which no such barrier is imposed by the correlation.

B. Photochemical Reactions

In photochemical reactions, initial electronic excitation is introduced by the absorption of a photon or by an energy transfer from another molecule. It is normally followed by a very rapid, radiationless conversion to the lowest excited singlet or the lowest triplet energy surface, depending on the multiplicity of the initial excited state. Also, any vibrational energy in excess of that dictated by the temperature of the surrounding medium, whether generated by the initial excitation or by the radiationless process, is rapidly lost to the solvent, unless one works in a gas phase at low pressure. Thus, in a matter of a few picoseconds or less the molecule ends up in one or another of the local minima in the S_1 or T_1 surface. Further motion on the surface may follow, depending on the temperature and the height of the barriers surrounding the local minimum. Also, radiationless conversion from the S_1 to the T_1 state, known as intersystem crossing, can occur. This often happens on a nanosecond time scale. Sooner or later a radiationless return to the S_0 state ensues. The final fate of the molecule is further loss of excess vibrational energy and thermal equilibration at the bottom of one or another catchment basin in the S_0 surface, depending on where on the S_0 surface the molecule landed. If this is the same minimum from which the initial excitation occurred, the process is viewed as photophysical. If it is not, a net chemical reaction has occurred and the process is labeled photochemical.

At times the excited S_1 surface may touch or nearly touch the S_0 surface, in which case the return to S_0 is very fast. Such areas in S_1 are often referred to as funnels since they very effectively return molecules to the ground state.

In order to understand photochemical reaction paths it is thus important to have an understanding of the location of barriers as well as minima and funnels in the S_1 and T_1 surfaces, plus a sufficient understanding of the S_0 surface to allow a prediction or rationalization of the fate of a molecule that lands in a known region of this surface. Correlation diagrams are often useful for this purpose. For instance, the diagram for the face-to-face cycloaddition of two ethylenes shown in Fig. 10 shows the presence of a minimum in the S_1 surface in the general area of geometries at which the pericyclic transition state occurred in the ground state. While the latter was energetically unfavorable in the S_0 state, making the reaction highly unlikely since the molecules will probably find other reaction paths, the minimum in the S_1 state provides an efficient driving force for the photochemical cycloaddition to proceed efficiently. Thus, reactions that fail to occur in the ground state are often smooth when performed photochemically and vice versa.

In general, by virtue of molecules landing at otherwise improbable and highly energetic areas on the S_0 surface, photochemical processes are capable of producing very highly energetic ground state products. Yet, frequently the same perturbations, such as substituent effects, that increase the stability of a molecule in the ground state also facilitate its photochemical reactions by lowering the barriers encountered along the way. The interplay of these two aspects of the excited-state surfaces—minima and barriers—make the consideration of photochemical processes far more complex than the study of thermal reactions.

SEE ALSO THE FOLLOWING ARTICLES

HYDROGEN BOND • INFRARED SPECTROSCOPY • KINETICS (CHEMISTRY) • MOLECULAR ELECTRONICS • ORGANIC CHEMISTRY, SYNTHESIS • PERIODIC TABLE (CHEMISTRY) • PHOTOCHEMISTRY, MOLECULAR • POTENTIAL ENERGY SURFACES • QUANTUM MECHANICS • RAMAN SPECTROSCOPY

BIBLIOGRAPHY

Albright, T. A., Burdett, J. K., and Whangbo, M.-H. (1985). "Orbital Interactions in Chemistry," Wiley, New York.

Burkert, U., and Allinger, N. L. (1982). "Molecular Mechanics," American Chemical Society Monograph, Washington, DC.

Čársky, P., and Urban, M. (1980). "*Ab Initio* Calculations," Springer–Verlag, Berlin.

Chemical Reviews, Special Issue on Reactive Intermediator **9**1(3) May 1991.

Dewar, M. J. S., and Dougherty, R. C. (1975). "The PMO Theory of Organic Chemistry," Plenum, New York.

Dewar, M. J. S. (1984). *J. Am. Chem. Soc*. **106,** 669.

Hehre, W. J., Radom, L., Schleyer, P. V. R., and Pople, J. A. (1986). "*Ab Initio* Molecular Orbital Theory," Wiley, New York.

Lowry, T. H., and Richardson, K. S. (1981). "Mechanism and Theory in Organic Chemistry," 2nd ed., Harper & Row, New York.

Michl, J., and Bonačić-Koutecký, V. (1990). "Electronic Aspects of Organic Photochemistry," Wiley, New York.

Salem, L. (1982). "Electrons in Chemical Reactions," Wiley, New York.

Schaefer, H. F., III, and Segal, G. A. (eds.) (1977). "Modern Theoretical Chemistry," Vols. **3, 4, 7,** and **8,** Plenum, New York.

Simons, J. (1983). "Energetic Principles of Chemical Reactions," Jones and Bartlett, Boston, MA.

Simons, J. (1991). An experimental chemist's guide to *ab initio* quantum chemistry. *J. Phys. Chem.* **95,** 1017.

Organic Chemistry, Synthesis

John Welch
State University of New York, Albany

GLOSSARY

Asymmetric synthesis Stereoselective preparation of a single enantiomer.
Biomimetic synthesis Construction of molecules by mimicking biosynthetic processes.
Diastereoselectivity Tendency of a reagent or reactant to form a single diastereomer in a reaction.
Electrophilicity Tendency of a reactant that is electron deficient to satisfy this deficiency in a reaction.
Enantioselectivity Tendency of a reagent to form a single enantiomer in a reaction.
Functional group Structural feature of a larger molecule which imparts to that molecule a particular chemical reactivity. Functional groups often contain heteroatoms but may also result from changes in hybridization or bonding.
Natural products Molecules formed by biological organisms. Natural products are frequently chiral and often complex.
Nucleophilicity Tendency of a reactant that has an electron pair available for bonding to contribute this electron pair in a reaction.
Regioselectivity Ability of a reactant to discriminate and principally form a single regioisomer.
Retrosynthetic analysis Systematic bond disconnection of a synthetic target for purposes of planning a synthesis.

ORGANIC SYNTHESIS, the science of the preparation of organic molecules, is a crucial element of organic chemistry. It is synthesis that makes possible the preparation of materials with novel structures and facilitates the study of naturally occurring substances otherwise available only after tedious and exacting isolation. The synthetic chemist is often called upon to test the latest advances in theory by preparing new molecules; yet the observations of the synthetic chemist frequently contribute to the development of new theories.

Organic chemistry, as the chemistry of carbon compounds, requires the functionalization of these compounds to prepare new substances. Once functionalized, construction of new carbon molecular frameworks is possible. Historically, the preparation of natural products has been the most important application of synthetic methods. Contemporary chemists not only seek to prepare the chiral natural compounds efficiently by asymmetric synthesis but, by utilizing biomimetic strategies, to copy the biosynthetic pathways.

Limitations of space restrict this article to highlights from this diverse field. Throughout the discussion, name reactions and procedures are cited in parentheses.

I. FUNCTIONAL GROUP MANIPULATION

As discussed in the introductory section, a key element of directed organic synthesis is the selective introduction and modification of functional groups. Manipulation of a functional unit may provide discrimination in chemical reactivity. Functional groups typically contain heteroatoms such as halides, oxygen as in alcohols or carbonyls, or nitrogen as in amines. Variation of the hybridization and bonding of carbon as in alkenes, alkynes, and arenes also constitutes functionality. Recently, carbon bonds to most of the elements of the periodic chart have been exploited by organic chemists; however, the chemistry of most of these more esoteric molecules is illustrated by those groups discussed earlier.

A. Halogenation

Halogenation is one of the most effective means to introduce functionality in saturated or unsaturated hydrocarbons. The relatively low reactivity of alkanes and alkenes requires reactive reagents for substitution. A large number of well-established methods for the interconversion of organic halides to other functionalities increases the utility of halogenated hydrocarbons as intermediates in the introduction of other functional groups. Conversely selective preparation of halides from alcohols carbonyls, or amines is also an important process.

1. Reactions of Alkanes and Alkenes

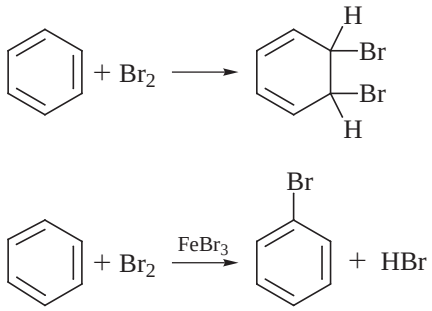
Reaction of an alkane with a halogen usually proceeds via a free radical pathway. The reactions of chlorine and bromine are more useful than those of fluorine or iodine. Direct fluorination of hydrocarbons with molecular fluorine is a very exothermic process resulting in complex product mixtures and extensive decomposition. Recently, it has been possible to improve the selectivity of the direct reaction of fluorine by dilution with an inert gas and by using very efficient cooling in a special apparatus.

Reactions with bromine or chlorine must be initiated either photochemically or by the use of a free radical initiator:

$$\begin{aligned} X_2 &\xrightarrow{h\nu} 2X\cdot \\ RH + X\cdot &\rightarrow R\cdot + HX \\ R\cdot + X_2 &\rightarrow RX + X\cdot \\ R\cdot + X\cdot &\rightarrow RX \end{aligned} \qquad (1)$$

As with fluorination, a mixture of products is formed. Photochemically induced iodination suffers from side reactions resulting from the hydrogen iodide formed.

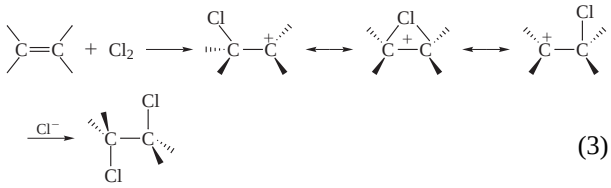
Although arenes will undergo radical addition reactions with loss of aromaticity, the Lewis acid promoted electrophilic halogenation of arenes is a substitution reaction where aromaticity is not lost.



(2)

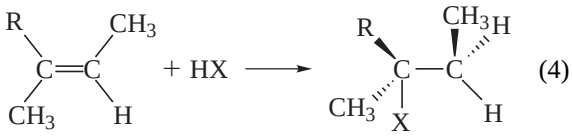
As an electrophilic substitution reaction, halogenation is governed by directing effects and can be regioselective.

In contrast to alkanes, alkenes readily add bromine and chlorine in electrophilic addition reactions:



(3)

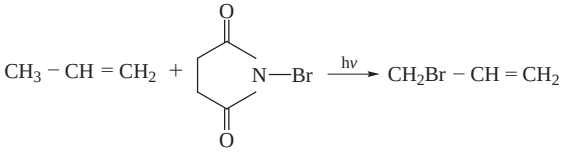
Hydrogen halides also add to alkenes following Markovnikov's rule, where the halide attaches to the more highly substituted carbon:



(4)

Hydrogen halide addition is occasionally accompanied by side reactions resulting from cationic rearrangements. Additions of hydrogen halides to alkynes are generally slow, forming, after addition of 2 moles of hydrogen halide, the geminal dihalides.

Alkenes may be treated with reagents such as *N*-bromosuccinimide in the presence of an initiator such as peroxide or light:

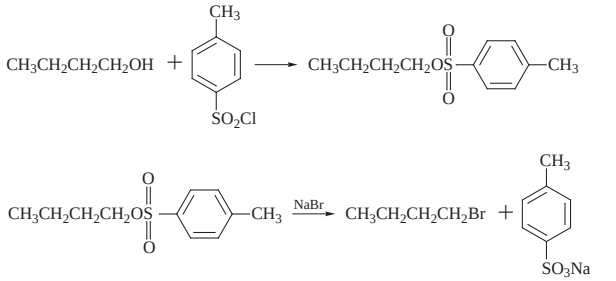


(5)

Under these conditions, where products of free radical attack are favored, the reaction can be very specific for allylic halogenation.

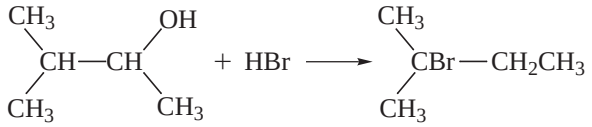
2. Reactions of Alcohols

Alcohols are easily converted to halides by displacement reactions. It is necessary to improve the leaving group ability of the hydroxyl by protonation or by conversion to a sulfonate or phosphate ester for successful displacement:



(6)

In reactions of primary and secondary alcohols with hydrogen fluoride, hydrogen chloride, or hydrogen bromide, competing rearrangements of the cationic intermediate may lessen the selectivity of the reaction:

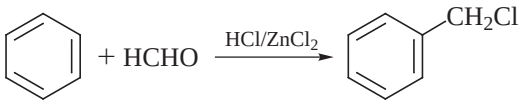


(7)

Treatment of an alcohol with thionyl chloride, sulfur tetrafluoride, phosphorous tribromide, or phosphorous pentachloride can yield the unrearranged halide regioselectively.

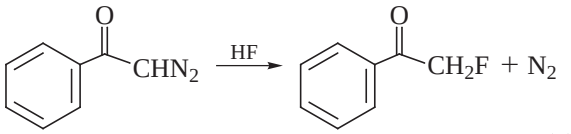
3. Haloalkylation Reactions

Halogenation can also be effected simultaneously with carbon–carbon bond forming reactions employing halogenated reactants. The chloromethylation reaction of aromatics with chloromethyl methyl ether and a Lewis acid is probably the best known example of such a reaction. It is possible to avoid the use of the toxic chloromethyl methyl ether by *in situ* formation of the reagent with formaldehyde and hydrogen chloride:



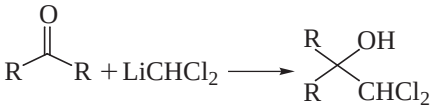
(8)

Alpha haloketones may be prepared by the reaction of carboxylic acid halides with diazomethane. Direct treatment of a diazoketone with a hydrogen halide, in a related reaction, also yields alpha haloketones or, under modified conditions, dihaloketones substituted with different halogens:



(9)

Dihalomethylation may also be effected by reaction of dihalocarbanions with carbonyl compounds or by the alkylation of a carbanion with chloroform:



(10)

Halomethylenation is possible by reaction of an appropriately halogenated Wittig reagent with an aldehyde or ketone.

B. Hydroxylation

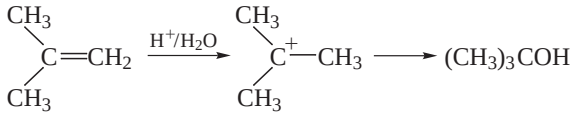
The selective direct hydroxylation of alkanes is not a practical general laboratory transformation although enzymatic hydroxylation of unactivated carbons is a major metabolic reaction in animals. Preparatively hydroxylation is more easily accomplished by hydrolysis of a halide. The ease of hydrolysis, iodide > bromide > chloride, parallels the synthetic utility of the reaction:

(11)

Hydrolysis of fluorides is of little consequene. In order to minimize side reactions, such as elimination to an alkene or rearrangement, S_N2 reaction conditions are desirable. However, under some conditions elimination reactions can be useful or desired. The synthesis of phenol from treatment of chlorobenzene with hydroxide ion first proceeds by elimination to form benzyne followed in a second step by the addition of water to form phenol.

1. Addition of Water to Alkenes

The direct hydroxylation of alkenes with aqueous acid results in the Markovnikov addition of water. Under these conditions, rearrangements of the intermediate cation are possible with a corresponding loss of regiospecificity.



(12)

Regiospecific addition of water in an anti-Markovnikov sense is readily possible via hydroboration of an olefin. The addition of borane or alkylborohydrides can be highly

regioselective. Oxidation of the boron–carbon bond with hydrogen peroxide generates the desired alcohol:

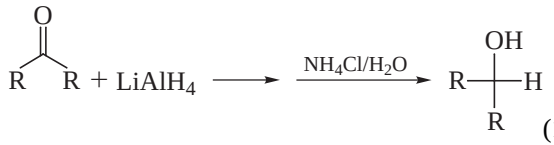
$$3CH_3CH{=}CH_2 + BH_3 \longrightarrow (CH_3CH_2C_2)_3B$$
$$\xrightarrow{H_2O_2/OH^-} 3CH_3CH_2CH_2OH \qquad (13)$$

When addition in a Markovnikov sense is required, oxymercuration can be very effective and is not accompanied by rearrangements. Addition of an electrophilic mercury species to an olefin, followed by trapping of the intermediate cation, leads to the formation of an oxygenated organomercurial. The carbon–mercury bond may be cleaved reductively with sodium borohydride:

$$CH_3CH_2CH{=}CH_2 + Hg(OAc)_2 \longrightarrow CH_3CH_2\underset{\displaystyle OAc}{\underset{|}{CH}}{-}CH_2HgOAc$$
$$\xrightarrow{NaBH_4} CH_3CH_2\underset{\displaystyle OH}{\underset{|}{CH}}{-}CH_3 + Hg^\circ \qquad (14)$$

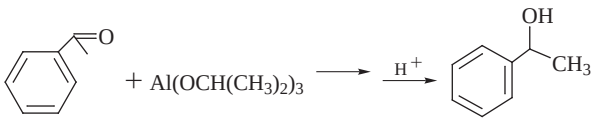
2. Reduction of Carbonyl Compounds

Aldehydes, ketones, carboxylic acids, and carboxylic acid derivatives can be reduced to alcohols. Historically, catalytic reduction with hydrogen and a catalyst such as platinum oxide or Raney nickel was employed to form the hydroxylic product in high yields. More recently, the more convenient metal hydride reducing agents such as lithium aluminum hydride or sodium borohydride and their derivatives have been employed to form alcohols at ambient or subambient temperatures:



(15)

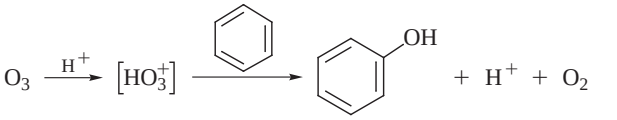
A variety of additional methods exist for the reduction of carbonyl compounds to alcohols. A typical example is the Meerwein–Pondorf–Verley reduction of carbonyls with aluminum alkoxides:



(16)

3. Electrophilic Oxygenation

Contemporary work has focused on the development of methods for the electrophilic oxygenation of alkanes, alkenes, and arenes:

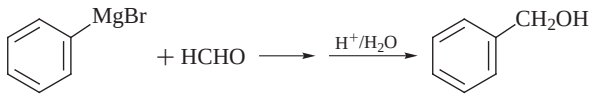


(17)

Treatment of ozone or hydrogen peroxide with very strong acids, such as hydrogen fluoride–antimony pentafluoride or fluorosulfuric acid–antimony pentafluoride, is proposed to lead to the formation of protonated ozone or hydrogen peroxide, which are highly electrophilic reagents.

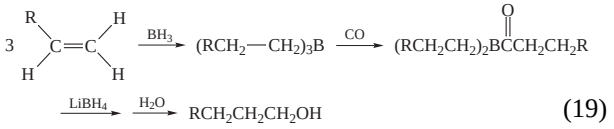
4. Hydroxymethylation

By simple analogy with haloalkylation chemistry, alcohols can also be introduced by carbon–carbon bond forming reactions. The addition of formaldehyde to a nucleophilic carbon such as the alpha carbon of an enol or organometallic reagent leads to hydroxymethylation:



(18)

In a different approach, treatment of an alkyl boron compound with carbon monoxide leads to carbon monoxide insertion in the alkyl carbon boron which can be reduced to form the hydroxymethyl group:



(19)

The Williamson ether synthesis, in which treatment of an alkyl halide with the alkali metal salt of an alcohol yields an ether, appends an alkoxy chain to an alkyl halide.

$$CH_3I + CH_3CH_2CH_2CH_2ONa$$
$$\longrightarrow CH_3CH_2CH_2CH_2OCH_3 + NaBr \qquad (20)$$

Alkoxymethylation is possible by reaction of substituted ethers with an appropriate nucleophile.

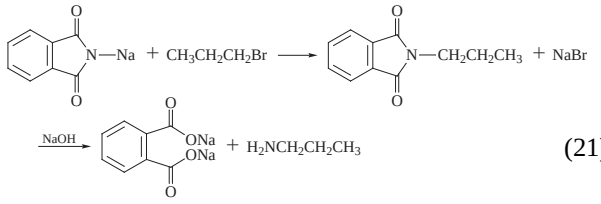
C. Amination

Selective formation of amines is extremely important for the synthesis of many natural products and biologically active compounds. The introduction of amines in the presence of other functionality is a particularly challenging synthetic problem. The selectivity required for practical utility is not generally associated with the more rigorous

conditions required for direct amination; therefore, functional group transformations are employed.

1. Displacement of Halides

The reaction of an alkyl halide with ammonia often leads to a mixture of primary, secondary, and tertiary amines where more than one alkyl halide molecule has reacted with a single ammonia. Selective formation of the primary amine may be possible if the alkyl halide is soluble in the presence of a large excess of ammonia. On a laboratory scale, the Gabriel synthesis or one of its modifications may be a more practical approach. The alkyl halide is treated with phthalimide or an alkali metal salt of phthalimide. The product alkylated phthalimide is hydrolyzed to selectively form the primary amine:

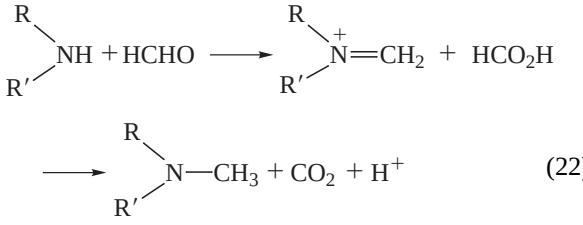


(21)

Selective formation of secondary or tertiary amines by reaction of an alkyl or dialkyl amine is generally ineffective for the same reason that the reaction of ammonia is not selective. However, direct reaction of an amine with an excess of an alkyl halide can be a useful method to the preparation of quaternary ammonium salts.

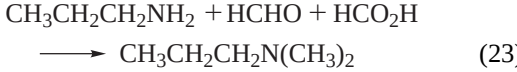
2. Derivatization of Carbonyl Compounds

An especially useful way to prepare secondary and tertiary amines is via an intermediate imine or iminium ion. For example, reaction of formaldehyde with a secondary amine in the presence of formic acid leads directly to a new tertiary amine:



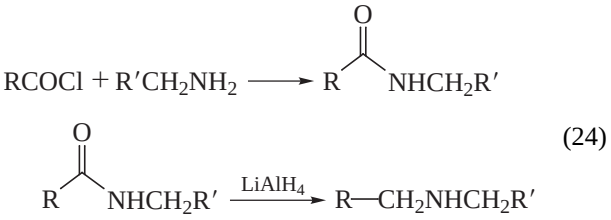
(22)

Dimethyl amines are conveniently prepared by reaction of formaldehyde with a primary amine in the presence of formic acid:



(23)

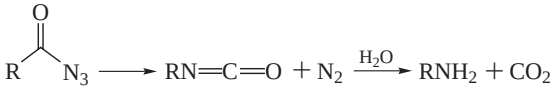
Acids may be converted to amides through reaction of the acid chloride with an amine:



(24)

Reduction of the acid amide to a new amine is possible by catalytic hydrogenation using a metal catalyst (such as platinum oxide, palladium oxide, palladium on barium sulfate, or Raney nickel). Metal hydrides such as lithium aluminum hydride, aluminum trihydride(alane), or diborane will also conveniently reduce amides to amines.

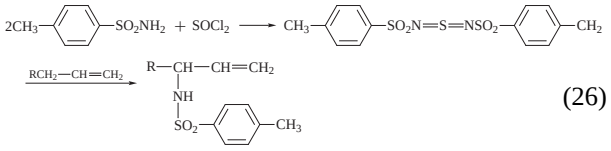
The direct conversion of acids to amines via an acyl nitrene intermediate is also possible by the Hofmann, Schmidt, or Curtius rearrangement:



(25)

3. Allylic Amination

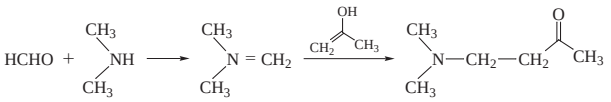
Allylic amination of alkenes via sulfur and selenium reagents is also possible. In the latter case, the reagent is prepared by treatment of selenium with chloramine T, in the former, by reaction of thionyl chloride with *p*-toluenesufonamide:



(26)

4. Mannich Reaction

Condensation of formaldehyde with an amine forms an iminum ion which can act as an electrophile in further synthetic transformations. One result is the addition of a carbon bearing an amine to the nucleophile. Commonly the reaction is effected with enolizable carbonyl compounds; however, the reaction will work with any compound with an activated hydrogen:



(27)

The intermediate iminium ion is even electrophilic enough to react with arenes.

D. Carbonylation and Carboxylation

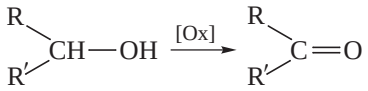
One of the most useful functional groups in organic synthesis is the carbonyl group. The carbonyl function reacts at the carbonyl carbon as an electrophile or may enolize and react at the alpha carbon as a nucleophile. Commensurate with the importance of carbonyl compounds to synthetic chemistry, a large number of methods have been developed for the preparation of these compounds.

1. Oxidation of Alcohols

A tremendous variety of oxidants are known for the conversion of alcohols to carbonyl containing compounds. The most difficult of these transformations is the oxidation of a primary alcohol to an aldehyde, the aldehyde often being more susceptible to oxidation, than the primary alcohol. Typical of the reagents employed to effect these oxidations are chromium and manganese oxides, dimethyl sulfoxide and dicyclohexylcarbodiimide (Pfitzner–Moffat), and dimethyl sulfoxide and oxalyl chloride (Swern):

$$RCH_2OH \xrightarrow{[Ox]} RCHO \qquad (28)$$

Oxidation of secondary alcohols to ketones is simpler if only because ketone products are more stable to further oxidation. Typically chromium reagents (CrO_3) or permanganate salts ($KMnO_4$) are used:



(29)

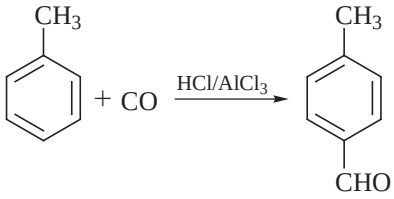
The conversion of primary alcohols or aldehydes to carboxylic acids is normally a facile reaction using $KMnO_4$. Aldehydes can be easily oxidized by air to acids, a result of the lability of the aldehydic hydrogen.

2. Carbonylation

The direct addition of a carbonyl group to olefins is promoted by metal catalysts such as rhodium, iridium, and cobalt. The oxo reaction, the commercial route to the preparation of aldehydes which are reduced to alcohols important as plasticizers, relies on such a carbonylation.

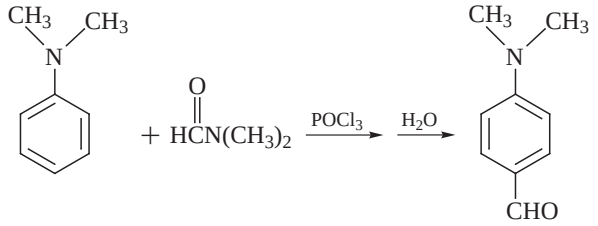
$$CH_2{=}CH_2 + CO + H_2 \xrightarrow{HCo[CO]_4} CH_3CH_2CHO \qquad (30)$$

The formylation of arenes by carbon monoxide and Lewis acids (Gatterman–Koch) is an important well-established method for the synthesis of arene aldehydes:



(31)

The use of carbon monoxide can be avoided by reaction of dimethylformamide in the presence of phosphorous oxychloride (Vilsmeier–Haack). The electrophilic intermediate chloroiminium ion adds to activated arenes. The substitution product is readily hydrolyzed to an aldehyde:



(32)

3. Nucleophilic Reagents

Carbonyl compounds may also be prepared by the reaction of a nucleophilic reagent such as an organolithium or Grignard reagent with formamide (Bouveault) of with an orthoester. In these examples, the initial product of reaction is a hemiaminal or acetal, respectively, which is unreactive to additional equivalents of the nucleophilic reagent but is readily hydrolyzed under acid conditions to reveal an aldehyde:

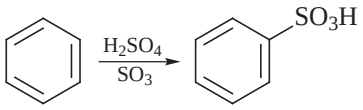
$$CH(OR')_3 + RMgBr \longrightarrow R{-}CH(OR')_2 + ROMgBr$$
$$RCH(OR)_2 + H^+/H_2O \longrightarrow RCHO \qquad (33)$$

The reactions are not limited to Grignard or organolithium reagents; any activated hydrogen compound will give similar results. In related reactions an organometallic compound will react directly with carbon dioxide to form on hydrolysis a carboxylic acid.

E. Introduction of Sulfur

The importance of sulfur in natural products and in products of commerce should be mentioned. Sulfur-containing compounds also have significant utility in directing further synthetic transformations.

Sulfonation of arenes is one of the oldest and best-known reactions for the introduction of sulfur into organic molecules. Treatment with sulfuric acid and oleum or chlorosulfonic acid leads to the preparation of sulfonated aromatics that are useful as surfactants, dye constituents, or chemical intermediates:



(34)

Sulfonation under acidic conditions is of little utility in the preparation of aliphatic sulfonates. Alkyl sulfonates are selectively prepared by reaction of alkanes with sulfur dioxide and chlorine. When the reaction is conducted in the presence of ultraviolet light, side reactions such as chlorination are effectively suppressed:

$$RCH_3 + SO_2 + Cl_2 \xrightarrow{h\nu} RCH_2SO_2Cl \quad (35)$$

The initial product of these reactions, as well as the product of the reaction of arenes with chlorosulfuric acid, are sulfonyl chlorides which may be hydrolyzed to the desired acids.

1. Sulfide Formation

Sulfides may be prepared by the reaction of alkali metal salts such as sodium sulfide with alkyl halides. Dialkyl sulfides are prepared by the reaction of the alkali metal salt of an alkyl sulfide with a second molecule of an alkyl halide:

$$RSNa + R'CH_2X \longrightarrow R'CH_2SR + NaX \quad (36)$$

Alternatively alkyl halides may be reacted with thiourea to form sulfides upon hydrolysis. Sulfides are also formed by the Markovnikov addition of hydrogen sulfide to alkenes. Alkyl sulfides will add in a conjugate manner to α,β-unsaturated carbonyl compounds, giving β-ketodialkyl sulfides.

II. CARBON–CARBON BOND FORMING REACTIONS

The preparation of the functional groups most important in organic synthesis has been described, but in the design of a synthesis, the construction of the carbon skeleton is often the greatest challenge. Frequently the desired functionality is either protected to prevent undesired side reactions or is carried through a synthetic sequence in a masked form, to be liberated after other transformations have been accomplished. These protected or masked functional groups have been described as synthons, structural units which can be formed or assembled by known or conceivable synthetic transforms. The retrosynthetic analysis of the carbon skeleton determines which carbon–carbon bonds should be formed and in what order they should be assembled. After analyzing the molecule and determining convenient fragments, the emphasis now focuses on determining which carbon–carbon bond forming reactions can be employed.

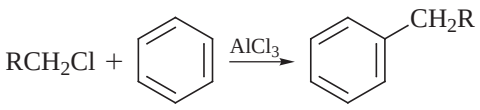
Carbon–carbon bond forming reactions can be described as being ionic processes, in which one fragment is electron deficient (electrophilic) and the other fragment is electron rich (nucleophilic), or as radical processes in which each fragment contributes a single electron. Traditionally, ionic processes have been the better understood and more effectively manipulated. However on occasion, reactions with radical intermediates have been described and effectively employed as well. Ionic processes will be described as resulting from nucleophilic or electrophilic components. Such a distinction is somewhat artificial, but parallels traditional descriptions for ionic carbon–carbon bond forming reactions in which one component must be electrophilic and the other nucleophilic. Typical electrophilic reactions such as alkylation and acylation and typical nucleophilic reagents such as organometallic compounds, enamines, and deprotonated imines will be discussed.

A. Alkylation

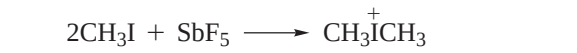
As mentioned earlier alkylations are described as electrophilic from the perspective of the alkylating agent. The electrophilic alkylation of alkanes or alkenes is known but is only of limited utility because of the difficulty of achieving selectivity. Nonetheless, the electrophilic alkylation of arenes is a reaction of substantial significance.

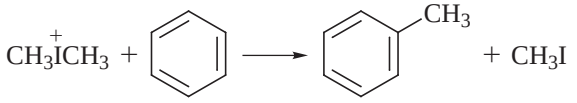
1. Alkyl Halides

Reports of the reaction of alkyl halides with arenes in the presence of Lewis acids appeared in the literature as early as 1877. Although alkyl fluorides are the most reactive of the alkyl halides, alkyl chlorides and bromides are the most widely used. Alkyl iodides are less commonly used as a result of accompanying side reactions and decomposition. Tertiary and benzylic halides are the most reactive, with secondary halides less reactive but more reactive than primary halides. Dialkyl halonium ions prepared by treatment of excess alkyl halide with antimony pentafluoride are especially reactive alkylating reagents. However, the most reactive alkylating agents are the methyl fluoride–antimony pentafluoride and ethyl fluoride–antimony pentafluoride complexes:

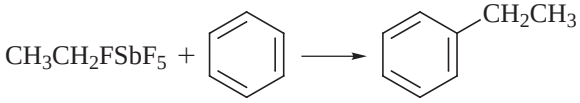


(37)





(38)



(39)

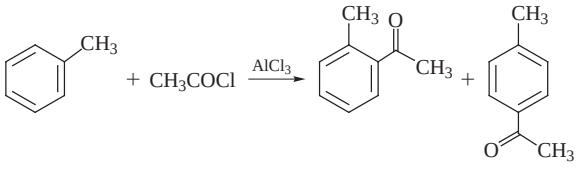
These powerful electrophiles will even react with poor nucleophiles such as alkanes.

2. Alkenes, Alcohols, Esters, and Ethers

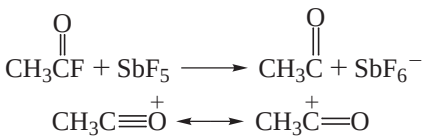
In addition to alkyl halides, alkenes will act as electrophiles in the presence of Lewis or Brønsted acids. Alcohols, esters, and ethers also will act as carbon electrophiles under acidic conditions.

B. Acylation

Acylation is another important electrophilic reaction for the substitution of arenes. The electrophilic reagent, an acyl cation, is stabilized by resonance. Frequently used as acylating reagents in the presence of a Lewis acid are acylchlorides, anhydrides, acids, and esters. Generally the reaction is performed by the addition of the Lewis acid catalyst to a mixture of the reactant arene and acyl halide:



(40)



(41)

Similar to the reactions of dialkyl halonium ions or methyl fluoride–antimony pentafluoride complexes, stable acylium ions, prepared by reaction of an acid fluoride with antimony pentafluoride, are extremely reactive acylating agents.

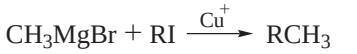
C. Alkylation of Organometallic Reagents

Organometallic reagents often react as carbanions and as such are nucleophilic. The electrophilic component of these reactions is generally aliphatic, but reactions of aromatic compounds are known. Although the alkylation of anionic carbon is well known, the mechanism of the reaction is still not clearly understood.

1. Alkyl Halides

The reactions of organometallic reagents with alkyl halides are coupling reactions. The reactions of alkyl chlorides, bromides, or iodides proceed well when the nucleophilic partner is a cuprate or a Grignard reagent (organomagnesium halide) in the presence of a copper, iron, or nickel catalyst:

(42)

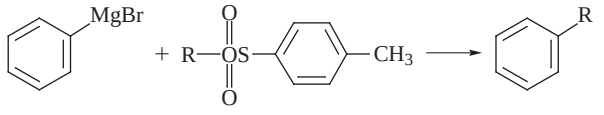


(43)

Allylic and propargylic halides also have been employed to alkylate organometallic reagents.

2. Alkyl Sulfonates

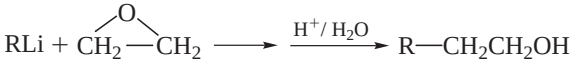
Often the alkylation of organometallic reagents is more effectively accomplished with an alkyl sulfonate. Sulfonates will undergo displacement reactions by both Grignard reagents and cuprates, frequently more cleanly and in better yield than the corresponding halides.



(44)

3. Epoxides

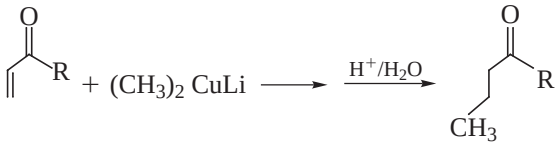
Epoxides may serve as a source of electrophilic carbon for reactions with organometallic reagents. The epoxide is opened, alkylating the organometallic while simultaneously forming a new alcohol.



(45)

4. α,β-Unsaturated Carbonyl Compounds

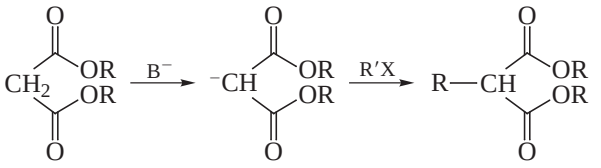
Nucleophilic organometallic reagents, particularly cuprates, will also add in a conjugate manner to α,β-unsaturated carbonyl compounds to form an enolate, stabilized by resonance. Bulky alkyl groups (R) enhance this reaction relation to carbonyl addition. Similarly, α,β-unsaturated sulfones will add organometallics to form a stabilized anionic product:



(46)

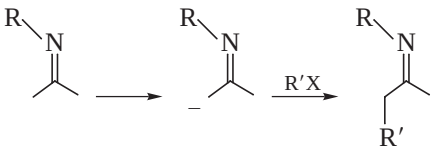
5. Alkylation of Active Hydrogen Compounds

Active hydrogen compounds are those in which the substituents present are capable of stabilizing the conjugate base formed on deprotonation of the starting material. Stabilization of the resultant anion significantly enhances the acidity of the proton, hence "active hydrogen" compounds. Common substituents which in this way stabilize negative charge are esters, aldehydes, ketones, acids, nitriles, nitro groups, sulfoxides, sulfones, sulfonamides, and sulfonates. Even further increased acidity is found on disubstitution, as in malonic esters. Deprotonation by even a weak base is sufficient to form the anion which may then be trapped by an alkyl halide. The reaction is a general one and applies to the other groups described as well:



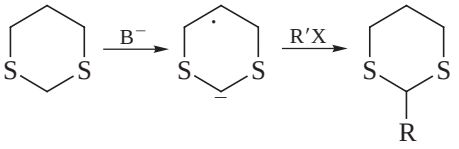
(47)

Imines, prepared by condensation of a primary amine with an aldehyde or ketone, may be deprotonated with a strong base to form the nitrogen equivalent of an enolate. The ready alkylation of these anions has found utility in enantioselective processes to be discussed later:



(48)

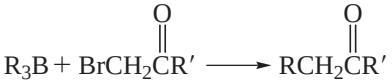
In addition to the oxidized forms of sulfur, dithioacetals, and ketals, such as 1,3-dithianes, also have activated hydrogens. Dithiane is readily deprotonated and easily alkylated. The aldehyde or ketone functionality masked by the dithioacetal or ketal is easily released under a variety of conditions:



(49)

6. Alkylation of Organoboron Reagents

Trialkyl boron reagents react readily with alpha halocarbonyl compounds, displacing the halide with an alkyl group. The reaction is of broad, general utility. Related reactions occur with diazoalkanes, esters and ketones:

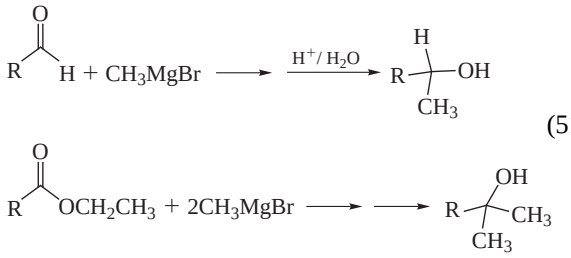


(50)

D. Reaction of Organometallic Reagents with Carbonyl Compounds

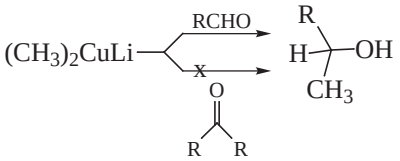
1. Grignard and Related Reagents

The addition of Grignard reagents to carbonyl compounds to form alcohols is an extraordinarily useful and general reaction. The reaction is facile and relatively simple. Ketones, aldehydes, and esters which require two equivalents of the Grignard reagent) react well:



(51)

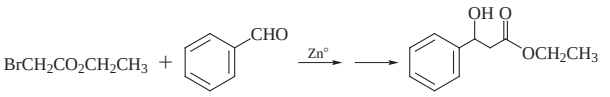
Other organometallic compounds, such as organolithium organozinc, and organoaluminum reagents, add to carbonyls. Noteably, organocadmium and organomercury reagents do not add. Selectivity is possible with organocuprate reagents which generally add to aldehydes but not to ketones:



(52)

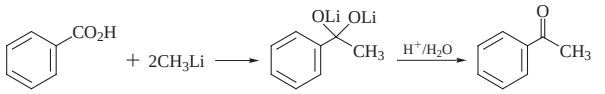
As mentioned earlier, addition of Grignard reagents to α,β-unsaturated carbonyl compounds often results in 1,4-rather than 1,2-addition reactions. In closely related

chemistry, α-bromo esters, acids, amides, and nitriles will add to carbonyl compounds in the presence of zinc dust. Presumably an organozinc reagent (Reformatsky), formed *in situ*, reacts like a Grignard reagent and adds to the carbonyl compound:



(53)

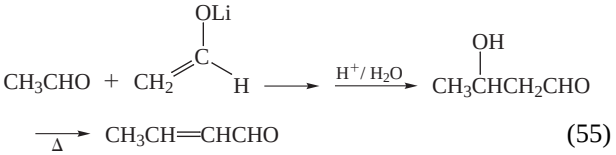
Organolithium reagents will add to the lithium salt of carboxylic acids to yield, on acidification, ketones:



(54)

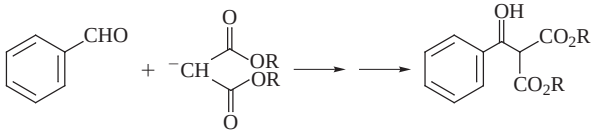
2. Addition of Active Hydrogen Compounds to Carbonyls

On deprotonation, active hydrogen compounds such as those substituted with carbonyl groups (as in esters, aldehydes, ketones, or acids) as well as nitriles, nitro groups, sulfoxides, sulfones, sulfonamides, and sulfonates will add to carbonyl compounds. The best known of these condensations is the aldol reaction. An enolate adds readily to ketones or aldehydes to form a beta hydroxy carbonyl compound:



(55)

Under more vigorous reaction conditions this product is readily dehydrated to give the α,β-unsaturated carbonyl compound. The reaction may involve self-condensation or may be a directed (crossed) aldol condensation between different carbonyl components (Claisen–Schmidt). An aldol condensation with a disubstituted enolate, e.g., malonate anion, is known as the Knoevenagel reaction:

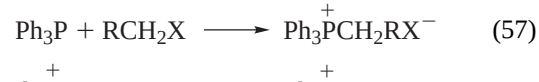


(56)

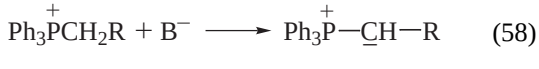
If substituted with an α-carboxylic acid, the aldol product is especially susceptible to decarboxylation.

3. Wittig Reactions

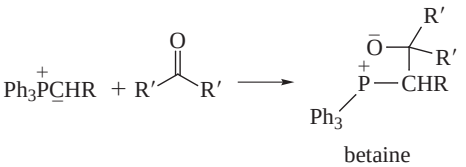
One of the most useful reactions of a stabilized anion with a carbonyl compound is the condensation of a phosphorous ylide with an aldehyde or ketone to form an olefin. The ylide is usually prepared from an alkyl triphenylphosphonium salt by deprotonation with a strong base. Addition of the ylide to the carbonyl compound initially forms a betaine intermediate which collapses to the olefin and triphenylphosphine oxide. The phosphonium salts are commonly prepared with triphenylphosphine and an alkyl halide; however, other less common phosphines have been employed as well:



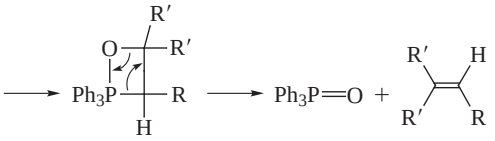
(57)



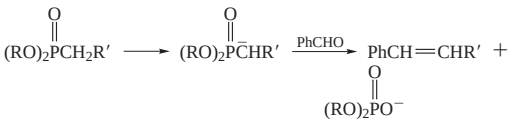
(58)



(59)



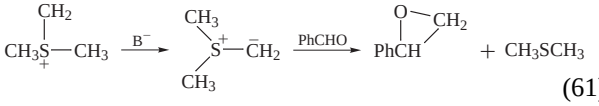
Ylides may also be prepared form phosphonates to yield a slightly more reactive ylide, in a variation known as the Horner–Emmons reaction:



(60)

The olefin product stereochemistry may be controlled by choice of reaction conditions to form selectively either the cis or trans product.

It is also possible to form reactive ylides with sulfur compounds. Typically, dimethyl sulfonium methylide will add to carbonyl compounds to form epoxides rather than alkenes:



(61)

III. NATURAL PRODUCT SYNTHESIS

The application of synthetic methods to the preparation of naturally occurring compounds has always been an important part of organic chemistry. From the early days

when separation and structural elucidation was the principal object, organic chemists have been interested in natural products. Rapidly, however, the synthesis of these materials came to be of importance, in part to verify proposed structural assignments and, of increasing importance today, to prepare quantities of natural materials with interesting pharmacological properties otherwise available only in very limited amounts. The preparation of natural compounds has also had a significant impact on the theory of organic chemistry, for example, the principle of the conservation of orbital symmetry may have resulted in part from insights developed during the synthesis of vitamin B-12. In this section, several examples have been taken from the synthesis of steroids and prostaglandins to illustrate the development of synthetic strategy.

A. Steroids

The early work of Robinson on the synthesis of the steroid nucleus may be contrasted with the elegant Woodward approach to the total synthesis of steroids. Robinson's approach requires the masterful manipulation of functionality so that through the course of the synthesis, relay compounds may be reached. Relay compounds are materials reached by alternate pathways, frequently by the degradation of a readily available natural product. The Robinson synthesis of androsterone, which spanned several years, required four stages, each stage comprised of numerous steps. The first stage was the transformation of 2,5-dihydroxynapthalene to the Reich diketone. The second stage was the transformation of the Reich diketone to the Koester and Logemann (KL) ketone. The third stage was the transformation of the KL ketone to dimethyl aetioallobilianate benzoate. The fourth required the conversion of dimethyl aetioallobilianate benzoate to androsterone. Both the KL ketone and aetioallobilianic acid were available from natural materials (Fig. 1).

In contrast to the lengthy manipulation of functionality to facilitate relay synthesis, Woodward's synthesis of methyl *d*,*l*-3-keto-Δ4,9(11). 16-etiocholatrienoate is considerably shorter and illustrates an advance in synthetic strategy. This efficient approach relies strongly on an understanding of stereochemical relationships and the efficient choice of functionality (Fig. 2).

B. Prostaglandins

The prostaglandins are a closely related family of compounds discovered as early as 1930 but whose structure was not determined until the 1960s. These compounds had a variety of physiological effects, but were only available in very minute quantities. As such they were ideal targets for synthesis. Biosynthetic analysis has shown that fatty acids, in particular arachidonic acid, are the

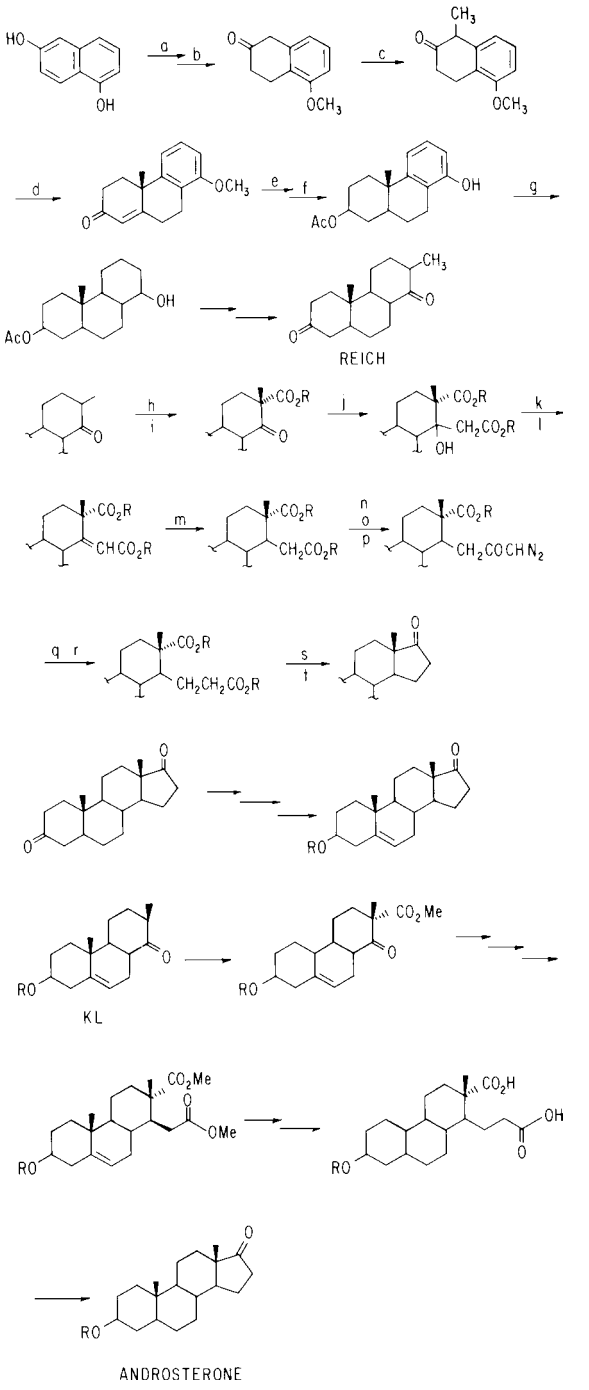


FIGURE 1 Robinson synthesis of androsterone. (a) Sodium methoxide, dimethyl sulfate; (b) sodium, ethanol; (c) sodium methoxide, methyl iodide; (d) diethylaminobutanone, methyl iodide, potassium; (e) hydrogen iodide; (f) platinum oxide, hydrogen; (g) palladium–strontium carbonate, hydrogen; (h) sodium triphenyl methide, carbon dioxide; (i) diazomethane; (j) ethyl bromoacetate, zinc; (k) platinum oxide, hydrogen; (l) phosphorous oxychloride; (m) platinum oxide, hydrogen; (n) potassium hydroxide, methanol; (o) oxalyl chloride; (p) diazomethane; (q) silver nitrate; (r) potassium hydroxide, methanol; (s) acetic anhydride; (t) heat.

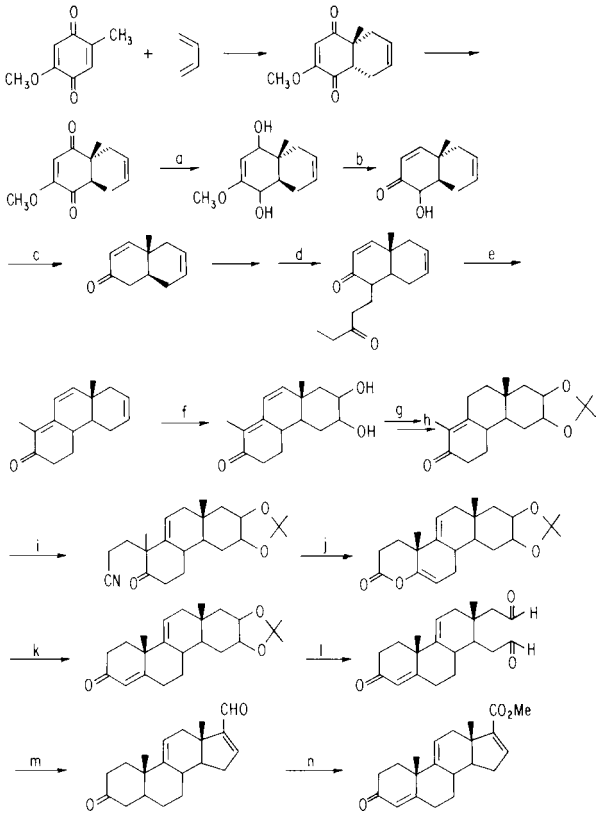


FIGURE 2 Woodward synthesis of methyl *d,l*-3-keto$\Delta^{4.9(11).16}$-etiocholatrienoate; (a) Lithium aluminum hydride; (b) acid, dioxane; (c) acetic anhydride, zinc; (d) ethyl vinyl ketone, potassium *t*-butoxide; (e) base, dioxane; (f) osmium tetroxide; (g) acid, acetone; (h) palladium–strontium carbonate, hydrogen; (i) acrylonitrile, Triton-B; (j) base; (k) methyl magnesium bromide, base; (l) periodic acid; (m) aqueous dioxane; (n) potassium dichromate, diazomethane.

precursors of the prostaglandins. A common feature of the prostaglandins was the cyclopentane ring, with as many as four adjacent stereocenters (Fig. 3). It was demonstrated by Corey and widely adopted by others that a common intermediate, "Corey's lactone," could be used to synthesize a number of the prostaglandins (Fig. 4). Prostaglandin synthesis from this common intermediate can therefore be convergent, the side chains being prepared by separate synthetic procedures and coupled intact to the cyclopentanoid system. This is in sharp contrast to both the Robinson and Woodward syntheses, which were highly linear, and demonstrates another advance in synthetic strategy.

Because of its utility, a number of syntheses for the Corey lactone have been developed. The variety of approaches which may be employed to reach this key intermediate demonstrates the power of convergent synthetic strategies. A few of the many routes to this compound are discussed in the following illustrations.

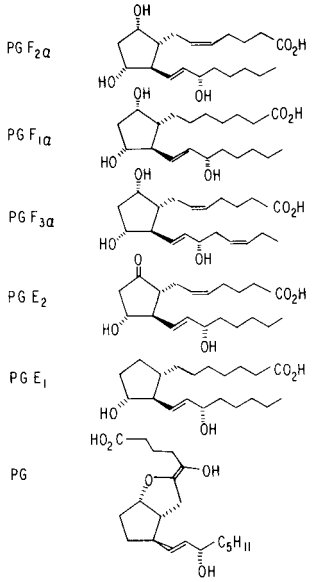


FIGURE 3 Prostaglandins.

A functionalized cyclopentadiene was allowed to react in a Diels–Alder reaction to yield a bicyclo-2.2.1-heptane molecule, which on further elaboration contained the correct relative stereochemistry and functionality (Fig. 5). The lactone was carried on to $PGF_{2\alpha}$ by Horner–Emmons reaction, by Wittig reaction, reduction, then deprotection (Fig. 6).

Bicyclo-2.2.1-heptadiene may also serve as a starting material. Under acidic conditions the functionalized nortricyclic intermediate was prepared. After a few additional manipulations the prescribed lactone was revealed (Fig. 7).

The lactol formed on reduction of the Corey lactone was also the target of several synthetic approaches. 1,3,5-Cyclohexanetriol was converted into the lactol via a ring contraction sequence. This approach could be made enantiospecific via resolution of an intermediate alcohol (Fig. 8).

Ring contraction was also the key step in the transformation of 1,3-cyclohexadiene to the lactol. An ene reaction

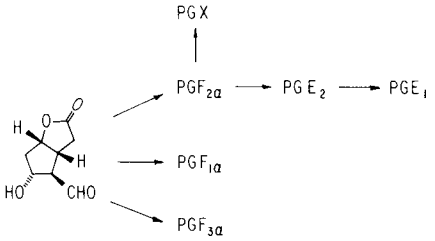


FIGURE 4 Synthetic tree demonstrating the utility of the Corey lactone in preparing prostaglandins.

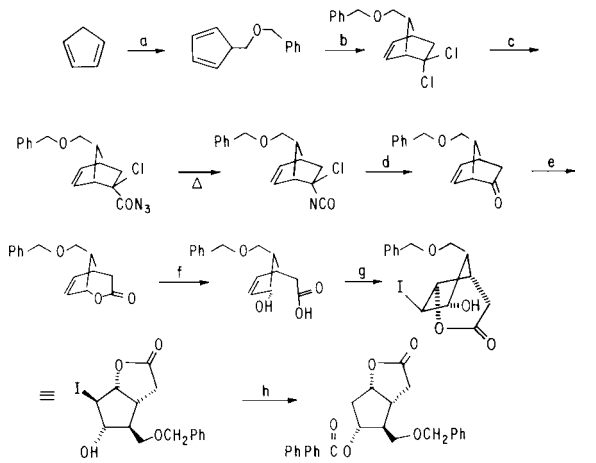


FIGURE 5 Synthesis of protected Corey lactone by cycloaddition to a functionalized cyclopentadiene. (a) Thallium sulfate, benzyl chloromethyl ether; (b) 2-chloroacryloyl chloride; (c) sodium azide; (d) aqueous acetic acid; (e) *m*-chloroperbenzoic acid; (f) base, carbon dioxide; (g) potassium iodide–iodine; (h) *p*-phenylbenzoyl chloride, tri-butyl tin hydride.

was employed in this synthetic scheme to introduce the appropriate substitution pattern (Fig. 9).

An asymmetric synthesis of the lactol from *S* malic acid required a homologation sequence followed by an intramolecular aldol condensation (Fig. 10).

IV. ASYMMETRIC SYNTHESIS

As was clear from the preceeding discussions of steroids and prostaglandins, natural products are often optically active. Their biological effects may be dependent on a specific configuration. Traditionally enantiomers were separated by resolution, i.e., the separation of the pair of diastereomers formed by reaction of the racemic product and an optically pure auxiliary. The use of resolution to

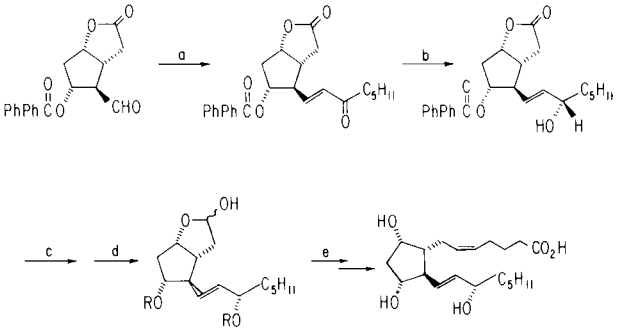


FIGURE 6 Elaboration of the Corey lactone to $PGF_{2\alpha}$. (a) $(CH_3O)_2POCHCOC_5H_{11}$; (b) lithium triethyl borohydride; (c) potassium carbonate, dimethoxyethane; (d) diisobutylaluminum hydride; (e) $Ph_3P{=}CH\ (CH_2)_3CO_2Na$.

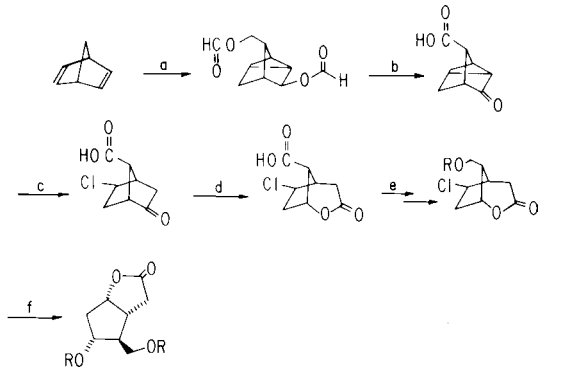


FIGURE 7 Bicyclo-2.2.1-heptadiene route to the Corey lactone. (a) Formaldehyde, formic acid; (b) chromic acid; (c) hydrochloric acid; (d) *m*-chloroperbenzoic acid; (e) ethyl chloroformate, zinc borohydride, dihydropyran; (f) base, hydrogen peroxide.

isolate the desired enantiomer can be feasible, but with the drawback that one-half of the material separated will be the undesired enantiomer as well as requiring an element of luck in the crystallization of the desired material.

Synthethc chemists have made remarkable strides in enantioselective reactions, where the desired enantiomer is the principle product. Asymmetric syntheses have successfully employed a variety of techniques, but we will limit discussion to enantioselective reducing agents, alkylations, and directed aldol reactions, which are typical of many other reactions and reagents.

For asymmetric synthesis it is necessary to employ a chiral component to direct the further transformations of the substrate. It is most desirable that this component be readily available and inexpensive, therefore common natural products such as terpenes or amino acids are frequently employed. The synthetic chemist may opt to recover the chiral auxiliary, in the case of reductions, or may chose to

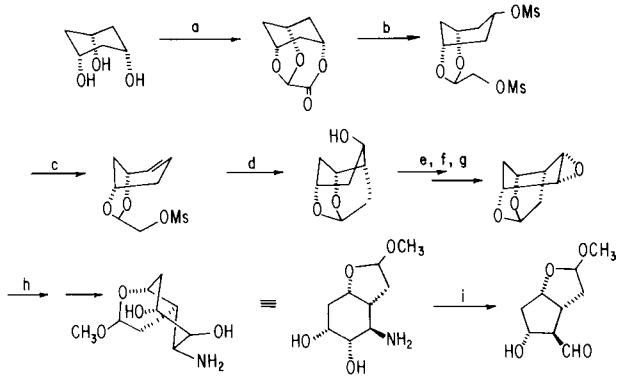


FIGURE 8 Woodward synthesis from 1,3,5-cyclohexanetriol. (a) Glyoxylic acid; (b) sodium borohydride, methanesulfonyl chloride; (c) potassium hydroxide; (d) potassium carbonate, dimethoxyethane; (e) methanesulfonyl chloride; (f) potassium hydroxide; (g) hydrogen peroxide; (h) aqueous ammonia, methanolic hydrogen chloride; (i) sodium nitrite, acetic acid.

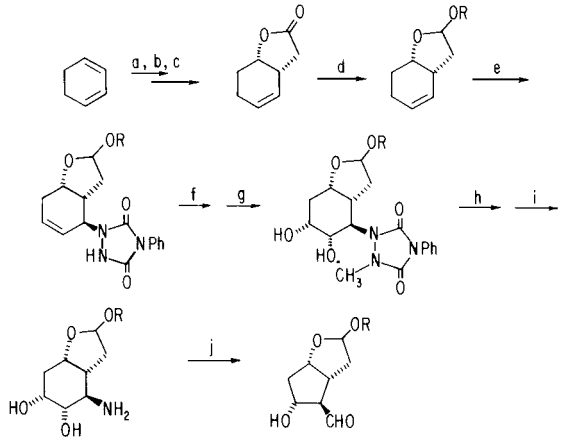


FIGURE 9 Preparation of prostaglandin precursor via an ene reaction. (a) Dichloroacetyl chloride; (b) zinc, acetic acid; (c) hydrogen peroxide; (d) diisobutylaluminum hydride; (e) *N*-phenyltriazolinedione; (f) sodium hydroxide, methyl iodide; (g) osmium tetroxide; (h) potassium hydroxide, methanol; (i) platinum oxide, hydrogen; (j) sodium nitrite, acetic acid.

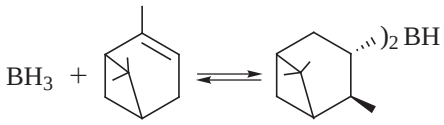
incorporate the chiral fragment in the carbon skeleton, as in directed aldol reactions.

A. Enantioselective Reductions

Among the most efficient and well-developed enantioselective transformations are asymmetric reductions of alkenes and carbonyl compounds.

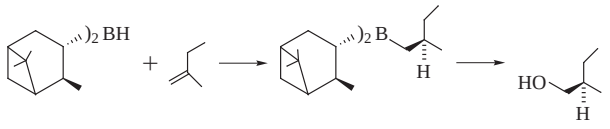
1. Hydroboration

Selective hydroborating reagents have been developed from readily available terpenes such as α-pinene and longifolene. The most useful reagents would react in high chemical and optical yields, yet permit recycling of the chiral auxiliary. One successful reagent is diisopinocampheyl borane, prepared by reaction of two equivalents of α-pinene with borane:



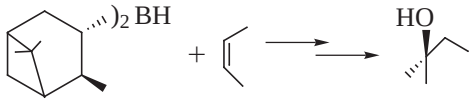
(62)

Reaction of the diisopinocampheyl borane with 2-methyl-1-alkenes, leads on oxidation to a disappointing 21% enantiomeric excess (i.e., the excess of one enantiomer relative to the other):



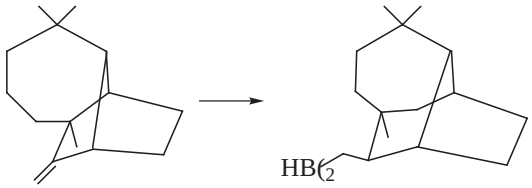
(63)

However treatment of *cis*-2-butene leads to e.e.'s as high as 98%. The general reaction of cis olefins appears to be limited only by the optical purity of the pinene starting material:



(64)

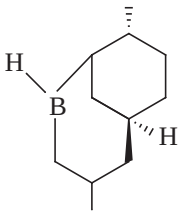
Dilongifolylborane, prepared by partial hydroboration of the longifolene, the most abundant sesquiterpene in the world, leads to only poor asymmetric induction in the reaction of 2-methyl-1-butene:



(65)

However, reaction of the cis alkenes is again more selective, e.e.'s as high as 78% are possible.

Limonylborane, a boraheterocycle with nonequivalent alkyl groups bound to boron, is prepared from limonene and has led only to disappointing 60% e.e.'s:



(66)

The most reactive of the hydroborating agents, monoisopinocampheyl borane reacts with cis olefins with 70% e.e. Also, the high reactivity of the reagent has led to its use with less reactive alkenes, with high e.e.'s having been reported:

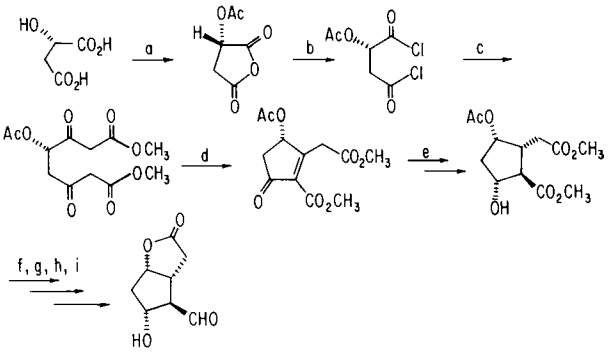
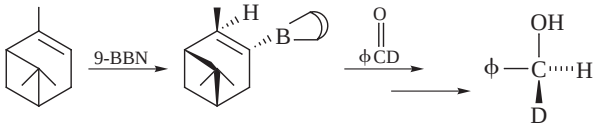


FIGURE 10 Chiral synthesis of the Corey lactone from malic acid. (a) Acetyl chloride; (b) dichloromethyl methyl ether, zinc chloride; (c) $HOOCCH_2CO_2CH_3$, base; (d) hydroxide; (e) hydrogenation; (f) potassium hydroxide, methanol; (g) acetic anhydride; (h) dichloromethyl methyl ether, zinc chloride: (i) sodium borohydride.

(67)

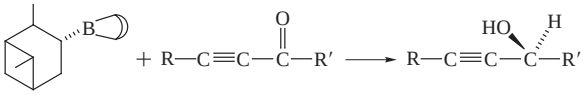
2. Chiral Borohydrides

The enantiospecific reduction of carbonyl compounds to alcohols is an extremely useful reaction. Aldehydes have been successfully reduced with chiral trialkylboranes to give products with very high optical purity. The hydroboration of α-pinene with 9-borabicyclononane (9-BBN) forms a chiral trialkylborane which reduces aldehydes with e.e.'s from 85 to 99%:



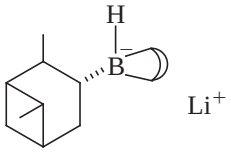
(68)

The reduction of normal ketones does not proceed well. However, the less sterically demanding acetylenic ketones can be reduced with 72–98% e.e.'s:



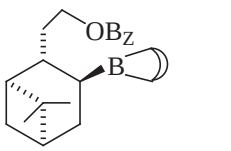
(69)

Complexed borohydrides have not yet lived up to their potential as enantiospecific reducing reagents. When the hydride reagent, lithium B-3-pinananyl-9-BBN-hydride, prepared by treatment of the α-pinene-9-BBN reagent with *t*-butyl lithium, was allowed to react with aldehydes the asymmetric induction was a disappointing 17–36%:



(70)

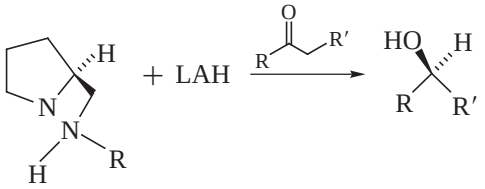
However, a related reagent prepared from nopol benzyl ether and 9-BBN gave e.e.'s as high as 70%. The presence of the benzyl ether side chain was predicted to improve asymmetric induction by improving coordination of the cation:



(71)

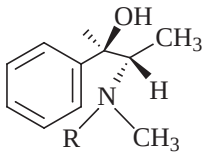
3. Complexed Lithium Aluminum Hydrides

Enantioselective reduction of aldehydes and ketones has also been possible with lithium aluminum hydride reagents complexed in a chiral environment. The lability of the ligands around lithium aluminum hydride reagents has so far limited the effectiveness of these reductions. The most effective ligands have been those containing at least one nitrogen. *N*-substituted amino methyl pyrrolidines, prepared from proline, have been used in the reduction of ketones to give optical yields as high as 96%:



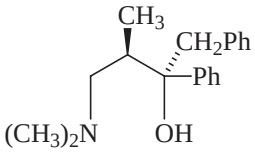
(72)

1,2-Amino-alcohols, such as ephedrine, also have been successfully used to reduce ketones with e.e.'s from 88 to 90%:



(73)

1,3-Amino-alcohols, such as Darvon alcohol, have been successfully employed in the enantioselective reduction of acetylenic ketones:



(74)

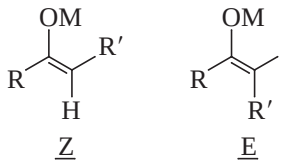
Although asymmetric reductions may be very useful procedures, it is important to recognize that the reductions are generally carried out at very low temperatures with excess reagents in order to maximize both the optical and chemical yields. These conditions often make these reagents prohibitively expensive for larger scale operations.

B. Asymmetric Syntheses with Enolates

The carbonyl group is one of the most useful functional groups because of its ability to act as an electrophile or, in the derived enolate, as a nucleophile. As described earlier, the enolate can react with alkylating agents or with carbonyl compounds in two very useful synthetic reactions. The stereoselectivity of an enolate is directly related to the stereochemistry of the enolate which is in turn affected by the stereochemistry of the parent molecule.

1. Alkylation Reactions

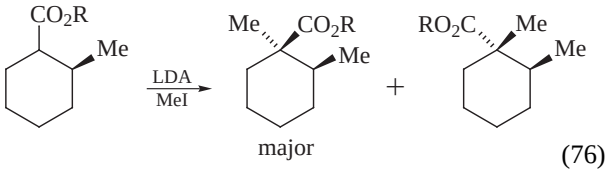
Stereoselective formation of the enolate is essential for stereoselective alkylation. An enolate may be either *E* or *Z* using the convention that the highest priority is always assigned to the OM group:



(75)

Enolate geometry may be controlled by careful choice of the deprotonation conditions. Use of sterically demanding lithium 2,2,6,6-tetra-methylpiperidide (LiTMP) favors formation of the *E* enolate. Addition of hexamethylphosphorous triamide (HMPT) to the reaction will reverse this selectivity to favor formation of the *Z* enolate. Use of the bulky base lithium hexamethyldisilazide (LiHMDS) will also favor formation of the *Z* enolate.

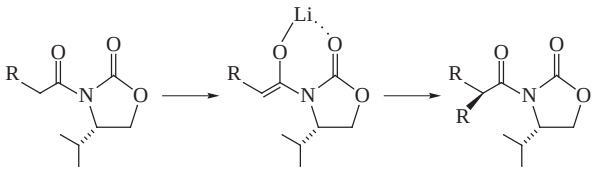
With control of the enolate geometry, in cyclic systems alkylation tends to follow the pathways which minimize steric interactions:



(76)

Employment of these interactions, with optically active functional groups, leads to control of the stereochemistry of alkylation.

In acyclic systems, even with control of enolate geometry, control of the alkylation reaction requires additional interactions. It is possible to create steric interactions comparable to those seen in cyclic systems by chelation effects with the enolate. If the chelating functionality is asymmetric, then stereochemical control will be possible. The result of such chelation is obstruction of one face of the enolate during the alkylation reactions:



(77)

A number of chiral auxiliaries derived from available chiral substrates such as valine, norephedrine, or proline have been successfully employed.

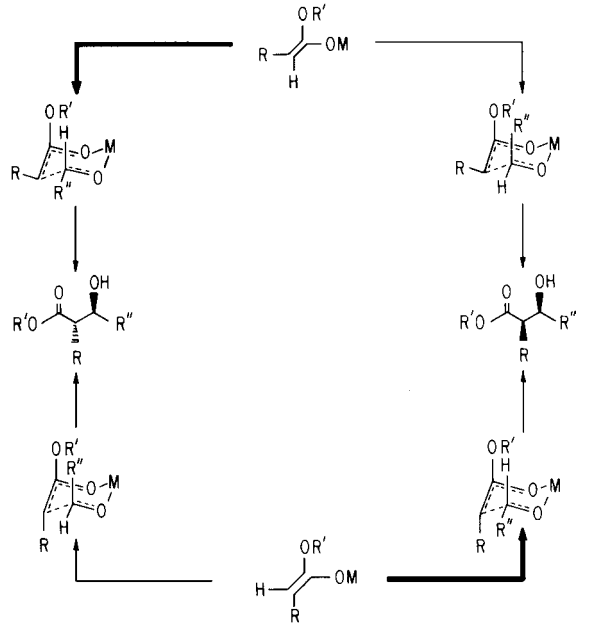


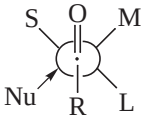
FIGURE 11 Zimmerman–Traxler transition state hypothesis for the directed aldol reaction.

2. Directed Aldol Reactions

The stereochemical selectivity of directed aldol reactions is also dependent on enolate geometry. Diastereoselectivity has been postulated to result from steric interactions in a six-membered cyclic transition state (Fig. 11). Diastereofaceselectivity results from addition to a chiral aldehyde, where the two new asymmetric centers may be controlled by a third center present in the starting aldehyde.

As illustrated in Fig. 11, as a general rule *Z* enolates tend to form syn aldols and *E* enolates tend to form anti aldols. The rule holds better for the reactions of *Z* enolates than *E* enolates. Selectivity increases with the steric demand of the enolate substituents. Diastereoselectivity can be poor even with good control of the enolate geometry if the steric interactions are not significant. Selectivity is also effected by the enolate counterion. The tighter the transition state, the more effective the steric repulsions will be in effecting diastereoselectivity. Metals with shorter bonds to oxygen, therefore, will be more effective. Boron enolates are generally more selective for a given carbonyl compound than other metal enolates.

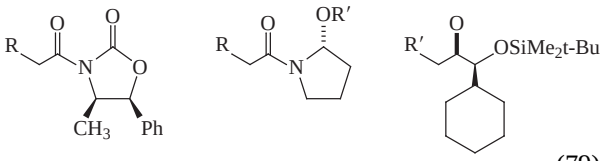
Diastereofaceselectivity describes the preference of an enolate to react with one face of a chiral aldehyde over the other. This problem has been analyzed many ways; however, Cram's rule is adequate in many cases. Cram's rule states that a nucleophile will react with a carbonyl from the same face as the smallest substituent when the carbonyl group is flanked by the largest substituent. Cram's rule is illustrated by a Newman projection below.



(78)

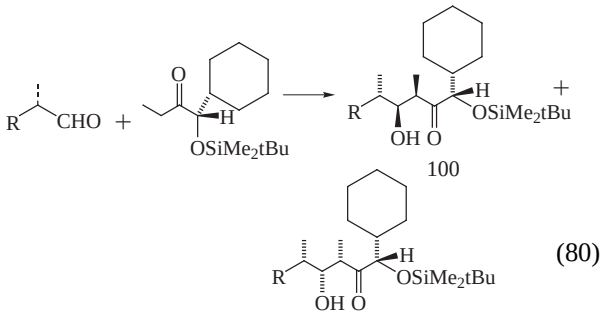
More complete analyses, taking into consideration the nature of the flanking substituents, which employ orbital overlap arguments, dipole effects, and chelation, have been advanced to explain results which cannot be understood by the simple Cram model. As might be deduced from this discussion, diastereofaceselectivity is remarkably dependent on the nature of the reactants.

Asymmetric induction is also possible in the reaction of chiral enolates with achiral aldehydes and ketones. This approach has been successfully employed with chiral imides, amides, and alpha hydroxyketones:



(79)

The asymmetric inductions realized with these chiral reagents may be amplified remarkably by their reaction with chiral aldehydes which have a complimentary asymmetry that tends to induce formation of the same enantiomer:



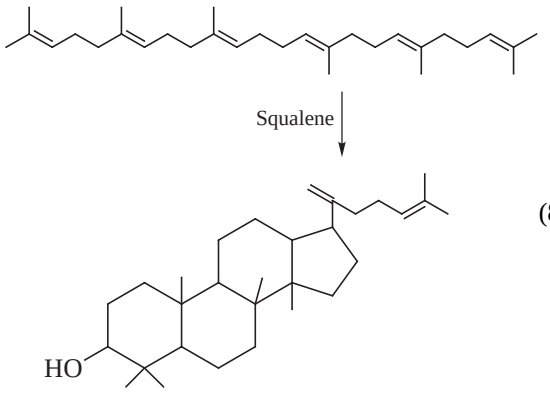
(80)

V. BIOMIMETIC SYNTHESIS

As discussed earlier natural products have stimulated the development of synthetic chemistry by providing challenging targets since the nineteenth century. In attempting to prepare known natural substances, it is only recently that chemists have turned to mimicking the synthetic approaches employed by nature. This discussion is limited to two different types of biomimetic synthesis, one based on the polyene cyclization reaction and the other the use of biomimetic reagents, functionalized cyclodextrins, as guest–host enzyme models.

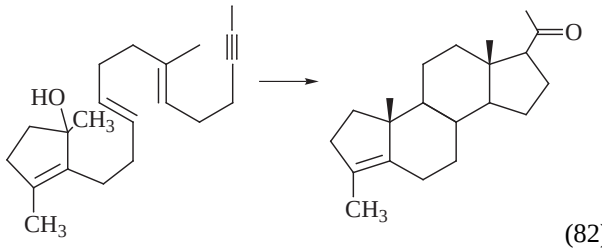
A. Biomimetic Polyene Cyclization

The independent proposal of groups at Columbia University and the ETH in Zurich that the stereoselectivity observed on the cyclization of polyenes is a consequence of the olefin stereochemistry constitutes the basis of the Stork–Eschenmoser hypothesis. The epoxide-initiated cyclization of squalene results in the stereospecific formation of dammaradienol. The all-trans double bond geometry of squalene results in the trans–anti-stereochemistry of the ring junctions in dammaradienol:



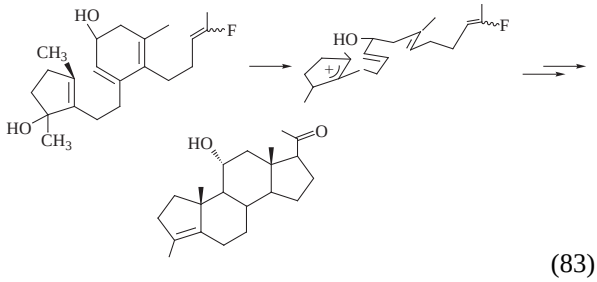
(81)

This postulate led W. S. Johnson to study the preparative utility of polyene cyclizations. The Johnson group studied numerous functional groups to both initiate and terminate cyclization of trans polyenes. The treatment of allylic alcohols under acidic conditions to form an allylic cation proved to be one of the more useful initiating functional reactions. Termination of cyclization by an alkyne, to form an intermediate vinyl cation which was then trapped, proved to be exceptionally efficient:



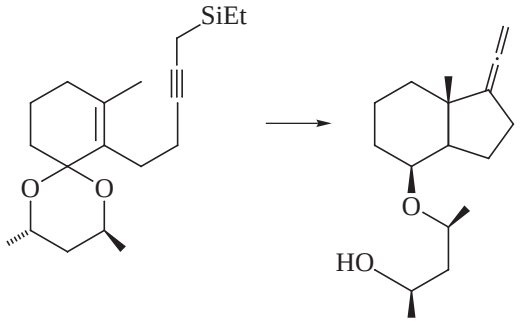
(82)

The efficiency of this process results in a 78% yield of tetracyclic material with very good control of the ring junction stereochemistry. This can be contrasted to the work of Robinson and Woodward described earlier. Particularly exciting was the observation that asymmetric cyclization was possible even when the asymmetric center was not involved in the bond forming process. Cyclization of a pro-C-11 hydroxy polyene with one of the best terminators, a vinyl fluoride, resulted in a 79.5% yield of the compound shown.



(83)

This material, with the correct configuration at C-11, may be efficiently converted to 11-hydroxy-progesterone, an intermediate in a commercial synthesis of hydrocortisone acetate. Johnson has shown that excellent asymmetric induction in polyene cyclizations is possible by treating acetals derived from (2*s*,4*s*)-pentanediol with Lewis acids:

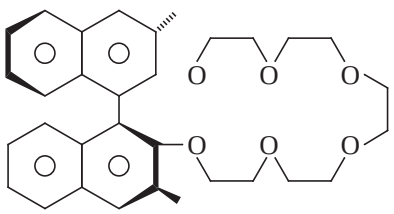


(84)

B. Guest–Host Inclusion Enzyme Models

Enzymes have unique catalytic, regulatory, and transport properties. The development of reagents that mimic these properties would be a significant synthetic advance. In early work it was possible to prepare model compounds that could duplicate simplistically some of the transport properties but with little substrate selectivity. Currently enzyme models are being prepared which not only have the ability to recognize the substrate but also can effect site specific reactions.

To promote substrate selectivity several types of inclusion host molecules have been developed. Remarkable successes in chiral recognition have been achieved with cyclic polyether compounds known as crown ethers:

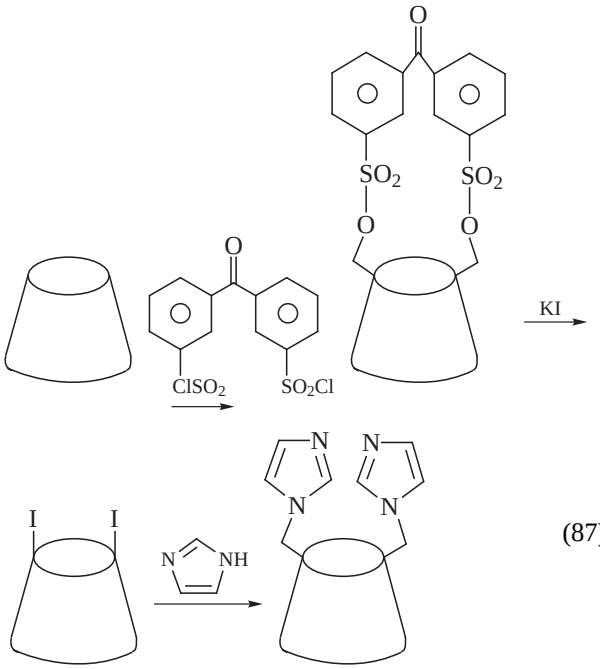


(85)

Cyclodextrins, oligosaccharides forming a distorted torus whose cavity may be occupied by other molecules, have also been explored as selective host molecules:

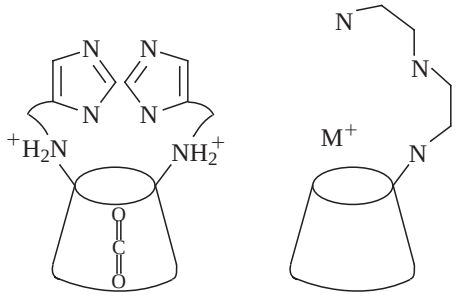
(86)

Cyclodextrins can have very regular cavity sizes. The cavity will incorporate aromatics to a uniform depth, often tilting the substrate away from the axis of the truncated conical cyclodextrin, most probably to maximize van der Waals contact. The limited size of the cavity results in substrate selectivity. It has also been possible to selectively functionalize cyclodextrins as illustrated by the preparation of a cyclodextrin substituted with imidazole units:



(87)

After the development of methods to prepare these molecules, the preparation of compounds more closely resembling enzymes was possible. Typical of these are cyclodextrins bearing a histamine-zinc complex and found to function as anhydrase models or polyamine metal complexes which act as carboxylic hydrolases:



(88)

There are numerous other examples of biomimetic synthesis. This narrow selection has been chosen to illustrate the progress and potential of this area.

SEE ALSO THE FOLLOWING ARTICLES

CATALYSIS, HOMOGENEOUS • ELECTRON TRANSFER REACTIONS • HALOGEN CHEMISTRY • HETEROCYCLIC CHEMISTRY • ORGANIC CHEMICAL SYSTEMS, THEORY • ORGANIC CHEMISTRY, COMPOUND DETECTION • ORGANIC MACROCYCLES • ORGANOMETALLIC CHEMISTRY • PHARMACEUTICALS • POLYMERS, SYNTHESIS • RUBBER, SYNTHETIC

BIBLIOGRAPHY

Boeckman, R. K., ed.-in-chief (1996). "Organic Synthesis," John Wiley & Sons, New York.

Burke, S. D., ed. (1999). "Handbook of Reagents for Organic Synthesis, Oxidizing and Reducing Agents," John Wiley & Sons, New York.

Coates, R. M., and Scott, E. D., eds. (1999). "Handbook of Reagents for Organic Synthesis, Reagents, Auxiliaries, and Catalysts for C─C Bond Formation," John Wiley & Sons, New York.

Diederich, F., and Stang, P. J., ed. (2000). "Templated Organic Synthesis," John Wiley & Sons, New York.

Fieser, M. (1999). "Fiesers' Reagents for Organic Syntheses," Vol. 18, John Wiley & Sons, New York.

Ho, T.-L. (2000). "Fiesers' Reagents for Organic Synthesis," Vol. 20, John Wiley & Sons, New York.

Lednicer, D. (1997). "Strategies for Organic Drug Synthesis and Design," John Wiley & Sons, New York.

Lednicer, D., and Mitscher, L. A. (1998). "The Organic Chemistry of Drug Synthesis," Vol. 6, John Wiley & Sons, New York.

Liska, F., and Volke, J. (1994). "Electrochemistry in Organic Synthesis," Springer-Veriag, Berlin/New York.

Mattay, J., ed. (1994). "Photochemical Key Steps in Organic Synthesis: An Experimental Course Book," John Wiley & Sons, New York.

Paquette, L. A., ed.-in-chief (1995). "Encyclopedia of Reagents for Organic Synthesis, Vol. 8, John Wiley & Sons, New York.

Pearson, A. J., ed. (1999). "Handbook of Reagents for Organic Synthesis, Activating Agents and Protecting Groups," John Wiley & Sons, New York.

The University of Liverpool, U.K. (2001). "Catalysts for Fine Chemicals Synthesis, Vol. 2, A Manual for Organic Synthesis," John Wiley & Sons, New York.

Ward, R. S. (1999). "Selectivity in Organic Synthesis," John Wiley & Sons, New York.

Zaragoza-Dorwald, F. (2000). "Organic Synthesis on Solid Phase: Supports, Linkers, Reactions," John Wiley & Sons, New York.

Organic Macrocycles

J. Ty Redd
Southern Utah University

Reed M. Izatt
Jerald S. Bradshaw
Brigham Young University

GLOSSARY

Binding constant Equilibrium constant associated with a reaction in which a ligand binds to a substrate.
Biomimetic That which mimics biological systems.
Calixarene Class of compounds consisting of a ring of phenol moities connected by methylene bridges.
Crown ether Cyclic organic compound consisting of a number of connecting ethylene oxide units.
Cryptand Organic compound consisting of two or more bridges connecting nitrogen atoms to give multiple rings.
Cyclic polyether See crown ether.
Cyclodextrin Class of cyclic polysaccharide compounds.
Enantiomeric recognition Differentiation of one enantiomer of the guest from the other by a chiral host.
Guest Chemical species that can be trapped by a host compound.
Host Compound that can trap a guest species.
Macrobicyclic effect Extra stability of complexes of macrobicyclic (cryptands) over those of crown ethers.
Macrocycle Cyclic ligand large enough to accommodate a substrate in the central cavity.
Macrocyclic effect Extra stability of complexes of cyclic ligands over those of analogous acyclic ligands.
Molecular recognition Ability of molecules to discern one another through molecular interactions that consequently form stable organized structures.
Self-assembly A spontaneous organization of molecules or objects into stable aggregates by noncovalent forces.
Spherand Type of macrocycle (see compound 11, Fig. 1).
Supramolecular chemistry A discipline that exploits fundamental concepts such as self-assembly, organization, and replication.

MACROCYCLES comprise a large group of heterocyclic organic compounds that can bind cationic, anionic, or neutral substrates by entrapment within the cavity created by the macrocyclic structure. The selectivity of the macrocycle for certain chemical species is a function of many parameters, an important one being the match between substrate size and macrocycle cavity size. When the substrate

is surrounded by the macrocycle structure, it is partly or completely isolated from the solvent. By this means, it is possible to solubilize bound substrates into solvents or membranes in which the unbound substrate is not soluble. Furthermore, the change in chemical reactivity of the bound substrate may be exploited to yield catalytic and biomimetic substrate transformation.

I. MACROCYCLIC STRUCTURE AND METAL CATION COMPLEXATION

Macrocycles, also called macrocyclic compounds, macrocyclic ligands, macropolycycles, and so forth, include a variety of basic structures. The major headings below group currently known macrocycles into broad, general classes. However, the term macrocycle is not confined to the limited number of representative compounds presented in this article.

A. Crown Ethers

The name *crown ether* was applied to cyclic polyether molecules such as compounds 1–3 in Fig.1, by Pederson, who first reported their preparation in 1967. A trivial nonrigorous nomenclature is commonly used to streamline naming of these complex molecules. Names are structured as follows: (1) principal ring substituents, (2) heteroatoms substituted for oxygen, (3) number of atoms in the principal ring, (4) the name *crown*, and (5) the number of heteroatoms in the principal ring. Thus, compound 1 is named 15-crown-5, compound 2 is 1,10-dithia-18-crown-6, and compound 3 is dibenzo-30-crown-10.

Crown ethers are particularly interesting ligands for two reasons: They measurably bind alkali metal cations in water solution, and they demonstrate size-based selectivity of metal ions. These features are illustrated for the ligands 15-crown-5, 18-crown-6, and 21-crown-7 in Fig. 2, where the thermodynamic equilibrium constant K for the reaction in methanol

$$M^{n+} + \text{18-crown-6} = (M - \text{18-crown-6})^{n+}$$

is plotted versus cation radius. Of the monovalent metal cations, K^+ is bound most strongly by 18-crown-6. X-ray crystallographic determination of the structure of the K^+–18-crown-6 complex shows that the K^+ ion sits at the center of the ligand cavity surrounded by the six ligand oxygen atoms as shown schematically in Fig. 3. The K^+ ion is nearly the correct size to fill the ligand cavity and is bound most strongly. The Na^+ ion is smaller than the 18-crown-6 ligand cavity, so the ligand must fold slightly to permit all six oxygens to associate with the cation. Both Rb^+ and Cs^+ are too large to fit into the ligand cavity. Thus, the relative sizes of cation and ligand cavity explain the selectivity of 18-crown-6 for K^+. Likewise, among alkaline earth cations, the Ba^{2+} ion both fits best in the 18-crown-6 ligand cavity and is bound most strongly. Table I, which lists the ionic radii of a number of metal cations, shows that K^+ and Ba^{2+} are of almost equal size.

Figure 2 shows that 21-crown-7, like 18-crown-6, binds the monovalent cation whose size matches that of the ligand cavity, that is, Cs^+. However, the selectivity of 15-crown-5 is not easily explained on the basis of relative size. While the cation Na^+ best matches the cavity in size, K^+ is bound slightly more strongly. This case illustrates that size is not the only factor and is often not the determining factor that controls cation selectivity. The problem arises from the fact that Na^+ ion is slightly too large to fit into the 15-crown-5 cavity. For such a case, cation solvation energies dominate in the free energy cycle

$$\begin{array}{ccc} M_g + L_g & \longrightarrow & ML_g \\ \Big| & & \Big| \\ M_s + L_s & \longrightarrow & ML_s \end{array}$$

for complex formation. (The subscripts g and s indicate species in the gas and solvent phases, respectively.) Compared to Na^+, the larger K^+ is less strongly solvated because of its lower charge-to-radius ratio, so less energy is expended in removing solvent molecules in the complexation process. Because the range of macrocycle sizes is much larger than the range of cation sizes, it is relatively rare that selectivity is governed by size predominantly. It is more often the case that solvation, ligand flexibility, and the effective charge on the binding sites play the dominant roles.

Crown ethers have affinity for metal ions besides those of the alkali and alkaline earth series. Figure 4 shows the binding constants of 18-crown-6 with the series of trivalent lanthanide cations, which decrease in size across the series. Table II shows the binding constants of several simple crown ethers with Pb^{2+}, Ag^+, Tl^+, and Hg^{2+}.

When the oxygen heteroatoms of crown ethers are replaced by nitrogen or sulfur, the selectivity of the ligands changes markedly. For example, sulfur-containing analogs of 18-crown-6 have lower affinity for alkali and alkaline earth cations and greater affinity for more polarizable cations such as Tl^+ and Hg^{2+}. When nitrogen is substituted, the affinity for alkali and alkaline earth cations also drops, while that for Pb^{2+} and Ag^{2+} increases.

Substitution of aliphatic or aromatic groups onto the heterocyclic backbone of crown ethers has a destabilizing effect on complex stability. Table III shows that the stabilities of complexes of dicyclohexano-18-crown-6 are much more like those of 18-crown-6 than are those of dibenzo-18-crown-6. In the latter case, the benzene rings withdraw electron density from the oxygen atoms, lowering the

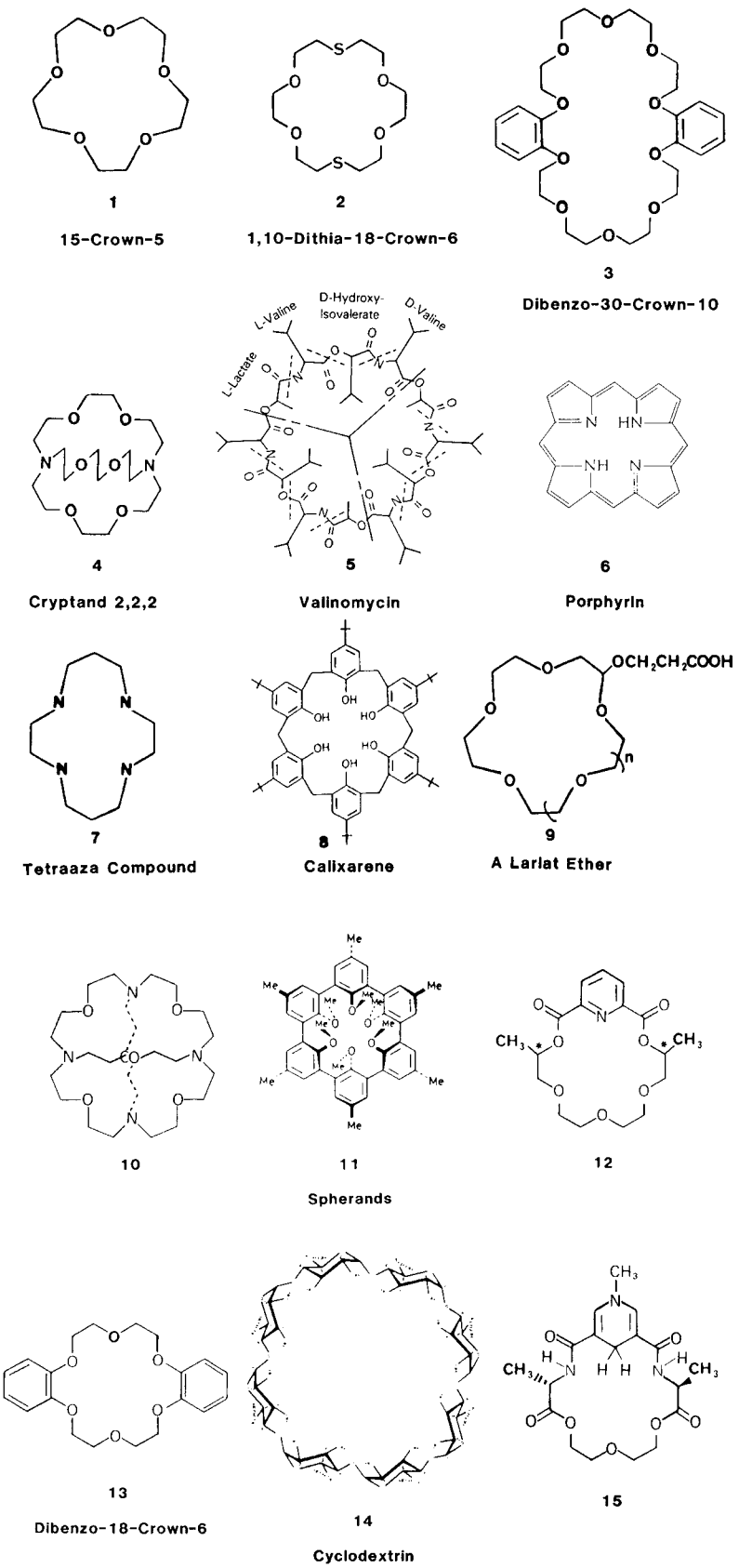


FIGURE 1 Structures of compounds discussed in this article.

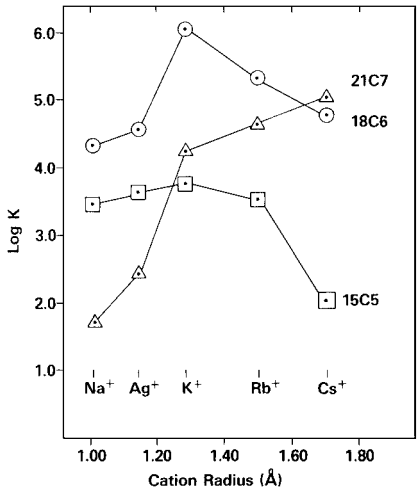


FIGURE 2 Log K for the reaction in CH_3OH at 25°C of several univalent cations with 15-crown-5, 18-crown-6, and 21-crown-6 versus cation radius.

energy of the ion–dipole interaction in the complex. This explanation is borne out in the observation that if electron-withdrawing substituents are added to the benzene rings, cation complex stability constants drop even further.

Bound metal ions show markedly different redox properties from unbound ions. It is possible, using macrocycles such as crown ethers, to stabilize oxidation states of metal ions such as Eu^{2+}.

B. Cryptands

Cryptands, or macrobicyclic ligands, are similar in structure to crown ethers, differing in the addition of a bridge that reaches across the ring to give football-shaped structures such as compound 4 in Fig. 1. The trivial nomenclature used for these ligands is given by the number of ethylene oxide units in each of the three bridges connecting

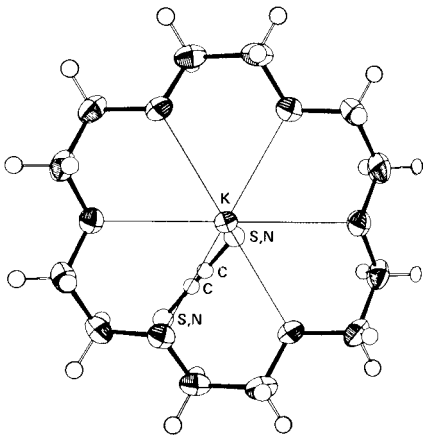


FIGURE 3 Structure of the potassium thiocyanate-18-crown-6 complex based on an X-ray crystallographic determination.

TABLE I Radii of Cations and of Crown Ether Cavities

Cation	Radius (Å)
Li^+	0.74
Na^+	1.02
K^+	1.38
Rb^+	1.49
Cs^+	1.70
Mg^{2+}	0.72
Ca^{2+}	1.00
Sr^{2+}	1.16
Ba^{2+}	1.36
Ag^+	1.15
Tl^+	1.50
Pb^{2+}	1.18
Hg^{2+}	1.02
Cd^{2+}	0.95
Macrocycle	**Radius (Å)**
15-crown-5	0.85
18-crown-6	1.3
21-crown-7	1.7

the two nitrogen heteroatoms. Thus compound 4 with two oxygen atoms in each bridge is designated 2.2.2. The additional bridge facilitates more effective encapsulation of metal ions. Consequently, the complexes of these ligands are in general more stable than those of crown ethers. Figure 5 shows the binding constants of a series of cryptands for alkali metal ions. Comparison of Fig. 5 with Fig. 2 shows the higher stability of the cryptand complexes. It also shows that the cryptands have a high degree of selectivity and that there is a cryptand of the correct size to be selective for each of the ions in the alkali metal sequence. The selectivity in all these cases is largely a result of the match between cation and cavity sizes. The effects of substituting other heteroatoms for oxygen and of organic substitution on the ring backbone are similar to those for crown ethers.

C. Naturally Occurring Macrocycles

Before crown ethers or cryptands were synthesized, macrocyclic ligands of various types had been isolated as natural products from microbial species. Compounds such as valinomycin and enniatin B have been studied extensively because of their ability selectively to bind alkali and alkaline earth metal ions and to transport such ions through biological membranes. Valinomycin is used in K^+ ion selective electrodes because of its high (10,000:1) $K^+ : Na^+$ selectivity.

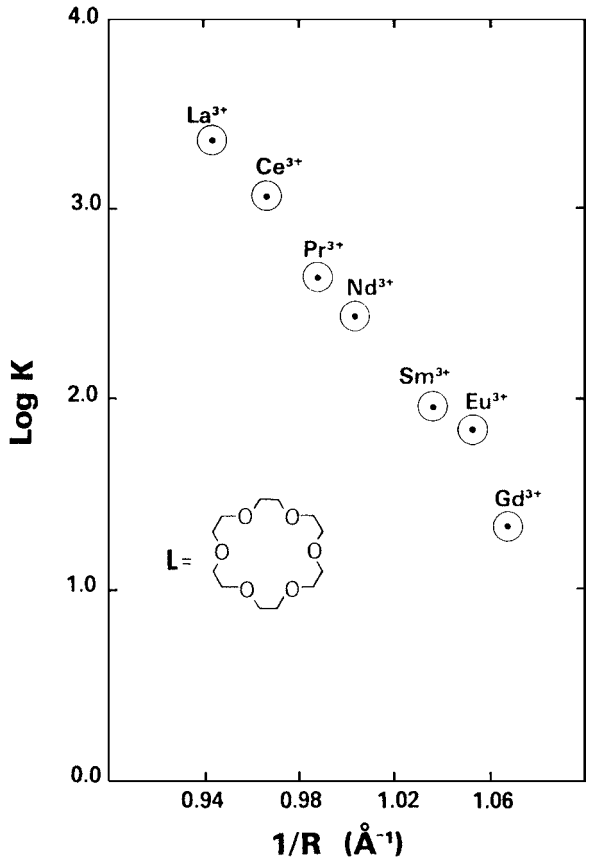


FIGURE 4 Log K for the reaction $M^{3+} + L = ML^{3+}$ (L = 18-crown-6) in methanol at 25°C versus reciprocal of cation radius. No reaction was observed with Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+}, or Lu^{3+}.

Valinomycin (compound 5, Fig. 1) is similar to a cyclic protein in structure. In the unfolded configuration, it is too large to accommodate metal ions. However, intramolecular hydrogen bonds cause the ring to tighten in on itself, providing a nearly octahedral arrangement of the oxygen atoms of the correct size to accommodate a K^+ ion. The exterior of the ligand is hydrophobic, making it soluble

TABLE II Stability Constants (log K) of Crown Ethers with Heavy Metal Ions in Water at 25°C

Ligand	Stability constant			
	Pb^{2+}	Hg^{2+}	Tl^+	Ag^+
15-Crown-6	1.85	1.68	1.23	0.94
18-Crown-6	4.27	2.42	2.27	1.50
Dibenzo-18-crown-6	1.89	—	1.50	1.41
Dicyclohexano-18-crown-6 (*cis*-anti-*cis*)	4.43	2.60	1.83	1.59
1,10-Dithia-18-crown-6	3.13	>5	0.93	4.34
1,10-Diaza-18-crown-6	6.90	17.85	—	7.8

TABLE III Effect of Substituent Groups on Cation Complex Stability (log K) with 18-Crown-6 and Its Analogs in Methanol at 25°C

Crown ether	Stability constant		
	Na^+	K^+	Ba^{2+}
18-Crown-6	4.36	6.06	7.04
Cyclohexano-18-crown-6	4.09	5.89	—
Dicyclohexano-18-crown-6 (*cis*-anti-*cis*)	3.68	5.38	—
Benzo-18-crown-6	4.35	5.05	5.35
Dibenzo-18-crown-6	4.36	5.00	4.28

in lipid membranes. Careful kinetic studies demonstrate that the selectivity of the ligand for K^+ is a function of the rate of cation release from the complex, there being little difference in the rate of cation uptake into the ligand.

D. Tetraaza Macrocycles

Macrocyclic ligands containing four nitrogen heteroatoms separated by various organic bridges have been studied

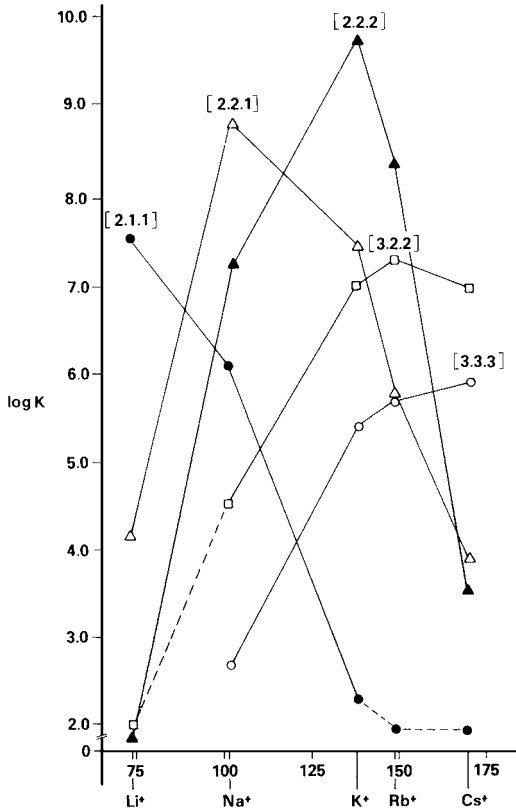


FIGURE 5 Variation of equilibrium constant K in 95 volume percent methanol for the reaction of several cryptands (see compound 4, Fig. 1) with the alkali metal ions Li^+–Cs^+ (plotted according to increasing metal ion radius).

for many decades. Examples are porphyrin (compound 6, Fig. 1) and its analogs, which are the metal-binding sites in many metalloenzymes, hemoglobin, and other naturally occurring compounds. There are also many synthetic tetraaza macrocycles that have affinity for divalent (and other) transition metal ions. Their binding constants with these cations are generally much higher than those typical of crown ethers and cryptands, so the metal ion is held virtually irreversibly. For example, the binding constant (log K) of compound 7, Fig. 1, with Hg^{2+} is 23 and with Ni^{2+} is 23.5. The importance of these complexes lies in the binding of additional ligands at the axial sites of the bound ion, which serves as the enzyme's active site. The degree and type of aromaticity in the tetraaza ligand structure has a profound influence on the electronic properties of the bound metal ion, which in turn affects the strength and nature of binding to additional ligands.

E. Other Macrocycles

A wide variety of macrocycle types has been reported in addition to the general categories discussed above. The calixarene ligands (compound 8, Fig. 1), which are water insoluble, have a strong, selective affinity for Cs^+. They form neutral complexes through loss of a proton. The lariat ethers (compound 9, Fig. 1) form neutral complexes by the same mechanism, resembling a crown ether with an arm that can reach around to provide ligation at the axial position. Macrotricyclic cryptands (compound 10, Fig. 1) provide essentially spherical or cylindrical neutral traps for metal ions. Spherands (compound 11, Fig. 1) likewise offer elegant binding geometries in which metal ions are bound. The list of macrocycles is far greater than can be presented in this limited space.

II. COMPLEXATION OF ORGANIC CATIONS

Ammonium and organosubstituted ammonium cations bind to crown ethers and other macrocycles by the formation of hydrogen bonds to the ligand heteroatoms. An example is the complex of an alkylammonium cation with 18-crown-6 shown in Fig. 6a. The stability of such complexes is influenced by the number of hydrogen bonds that can form and by the degree of steric hindrance for approach of the substrate to the ligand. Table IV lists the binding constants for a number of ammonium cations with 18-crown-6. The stability drops dramatically as the number of available hydrogen bonds is reduced from 3 to 2 to 1. Furthermore, anilinium ions, which contain ortho substituents, form weak or no complexes because the substituents sterically hinder approach of the $-NH_3^+$ group to the ligand.

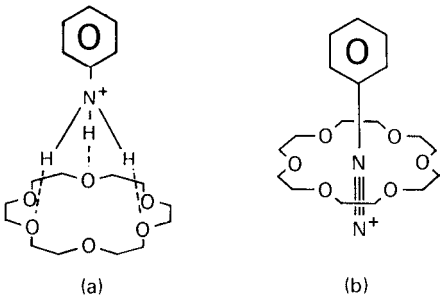


FIGURE 6 Diagrammatic representation of mode of binding of (a) anilinium and (b) benzenediazonium cations to 18-crown-6.

Unlike ammonium cations, diazonium cations complex to crown ethers by insertion of the positive moiety into the cavity, as in Fig. 6b. Table V shows that complex stability deteriorates markedly with ortho substitution in benzenediazonium cation because of steric hindrance. Figure 7

TABLE IV Stability Constants (log K) for Reaction of 18-Crown-6 and with Several Organic Ammonium Cations in Methanol at 25°C

Cation	Log K
RNH_3^+ cations	
NH_4^+	4.27 ± 0.02
$HONH_3^+$	3.99 ± 0.03
$NH_2NH_3^+$	4.21 ± 0.02
$CH_3NHNH_3^+$	3.41 ± 0.02
$CH_3NH_3^+$	4.25 ± 0.04
$CH_3CH_2NH_3^+$	3.99 ± 0.03
$CH_3CH_2OC(O)CH_2NH_3^+$	3.84 ± 0.04
$CH_3(CH_2)_2NH_3^+$	3.97 ± 0.07
$CH_3(CH_2)_2NH_3^+$	3.90 ± 0.04
$CH_2CHCH_2NH_3^+$	4.02 ± 0.03
$CHCCH_2NH_3^+$	4.13 ± 0.02
$(CH_3)_2CHNH_3^+$	3.56 ± 0.03
$CH_3CH_2OC(O)CH(CH_3)NH_3^+$	3.28 ± 0.02
$(CH_3)_3CNH_3^+$	2.90 ± 0.03
$PhCH(CH_3)NH_3^+$	3.84 ± 0.01
$PhNH_3^+$	3.80 ± 0.03
$2\text{-}CH_3C_6H_4NH_3^+$	2.86 ± 0.03
$4\text{-}CH_3C_6H_4NH_3^+$	3.82 ± 0.04
$2,6\text{-}(CH_3)_2C_6H_3NH_3^+$	2.00 ± 0.05
$3,5\text{-}(CH_3)_2C_6H_3NH_3^+$	3.74 ± 0.02
$R_2NH_2^+$	
$NH_2C(NH_2)NH_2^+$	21.7 ± 0.02
$(CH_3)_2NH_2^+$	1.76 ± 0.02
$(CH_3CH_2)_2NH_2^+$	
R_3NH^+ cations	
$(CH_3)_3NH^+$	No complex
R_4N^+ cations	
$(CH_3)_4N^+$	No complex

TABLE V Stability Constants (log K) for Reaction in Methanol at 25°C of Arenediazonium and Anilinium Cations with 18-Crown-6

Cation	Log K
$PhNH_3^+$	3.80
$2\text{-}CH_3C_6H_4NH_3^+$	2.86
$2,6\text{-}(CH_3)_2C_6H_3NH_3^+$	2.00
$PhNN^+$	2.50
$2\text{-}CH_3C_6H_4NN^+$	*a*
$2,6\text{-}(CH_3)_2C_6H_3NN^+$	*a*

[a] No measurable reaction.

shows that complex stability is a regular function of the electron density in the N_2^+ moiety. As electron density increases, stability declines.

The binding of organoammonium-type cations to macrocycles is the subject of intense interest due to the ability of such systems to mimic enzymes. Specifically, it is possible to add functionalities to the macrocyclic structure to permit chiral recognition in substrates. Chiral macrocycle 12 in Fig. 1, in the (S,S)-form, for example, binds one enantiomer of α-(1-naphthyl)ethylammonium perchlorate more strongly than the other isomer [$\log K = 2.47 \pm 0.01$ for the (R)-ammonium salt and 2.06 ± 0.01 for the (S)-salt]. The difference in binding constants results from a greater steric hindrance to the approach of the substrate for one isomer due to the presence of the bulky functionalities.

III. COMPLEXATION OF ANIONS AND NEUTRAL MOLECULES

Considerably less attention has been given to the binding of anions to macrocycles than to that of cations. Basically

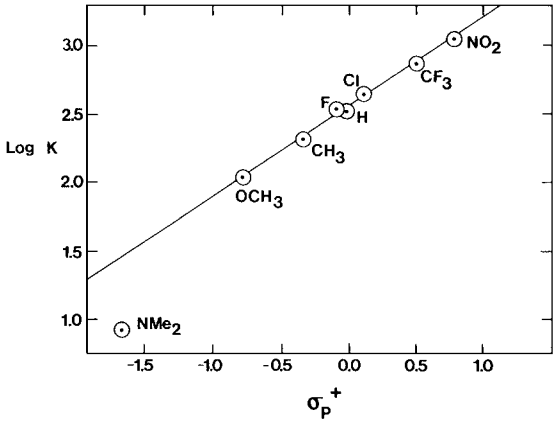


FIGURE 7 Plot of log K for formation in methanol at 25°C of the 18-crown-6 complex of $p\text{-}RC_6H_4NN^+$ versus Hammett σ_p^+ values of R.

TABLE VI Stability Constants (log K) for Reaction in Anion with Macrocyclic Ligand 10

Anion	Log K, reaction with ML 10 (protonated)
Cl^- (in water)	>4.0
Br^- (in water)	<1.0
Br^- (in 90% methanol)	1.75

two types of anion binding are known. The first involves macrocyclic structures containing basic sites that provide positively charged binding sites when protonated. Examples of these are compounds 10 (in protonated form) and 14 in Fig. 1, and binding constants with typical anions are found in Table VI. Structure 14 is based on cyclodextrin, which is a large cyclic polysaccharide.

Cyclodextrins are able to accommodate neutral molecules as well as anions. The large cavity contains numerous hydrogen-bonding sites if needed. In general, the cavity simply provides a comfortable microenvironment for many neutral species, especially when the solvent environment is less than ideal.

IV. APPLICATIONS OF MACROCYCLIC LIGANDS

A. Extractants

One of the first identified uses for crown ether and cryptand ligands was as phase transfer agents in catalyzing synthetic organic reactions. The macrocycle can be used to solubilize salts having oxidizing or reducing anions into hydrophobic solvents. For example, $KMnO_4$ can be solubilized into benzene by 18-crown-6. The "naked" MnO_4^- ion that accompanies the K^+–18-crown-6 complex in solution is a very powerful oxidizing agent in this medium. Use of naked anions of this type has provided a method to enhance the efficiency of many synthetic reactions.

Macrocycles have also been proposed as metal ion extractants in separation processes. It has been shown by J. McDowell and his co-workers at Oak Ridge National Laboratory that synergistic effects occur when crown ethers are used as coextractants with traditional extractants like diethyl hexylphosphoric acid (HDEHP). Specifically, the degree of metal ion extraction is greater when both crown and HDEHP are used together than the sum of extraction efficiencies when each is used separately. Furthermore, by using the crown, the selectivity of extraction processes can be altered in this manner.

B. Selective Ion Separations

Macrocyclic crown ethers can be attached to silica gel through stable C–Si and Si–O–Si bonds (see Fig. 8) by

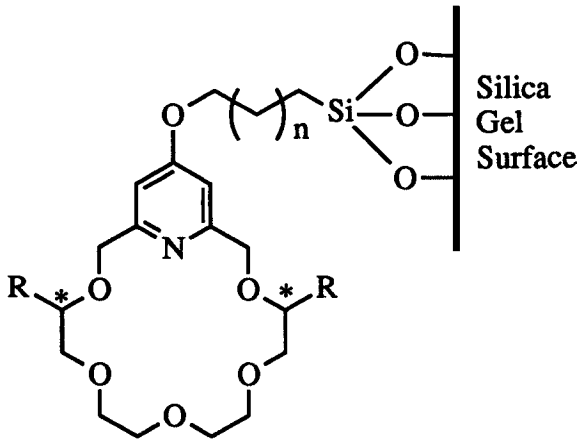


FIGURE 8 Chiral pyridino-18-crown-6 macrocyclic host attached to a solid support (silica gel).

heating a mixture of silica gel and crown ether-containing ethoxy silanes. These solid-supported macrocycles are currently finding success in the selective separation of metal cations. The macrocycle is permanently bound to the solid support and cannot be removed unless the silica gel is destroyed, as in concentrated aqueous base. This provides avenues for reusable separation systems.

Equilibrium constants for the association of metal cations with the silica gel-bound crown ethers are comparable to log K (H_2O) values for the association of the same cations with unbound ligands. These bound macrocycle systems can completely retain the desired metal ions while allowing the other metal ions to pass through the system. This method of using solid supported macrocyclic ligands to remove, separate, and concentrate specific metal ions from aqueous solutions has been commercialized for use in many analytical, environmental, metallurgical, precious metal, nuclear, and industrial operations. For example, this molecular recognition technology has been used in a pilot plant to separate and purify desired metal ions at a superfund site in Butte, Montana. The composition of metal contaminants in the water at this site includes Al, As, Cd, Ca, Cu, Fe, K, Mg, Mn, Ni, Na, Pb, Zn, and Si. The anions are chloride and sulfate. The pH of this solution is approximately 3 and the reduction potential is approximately 620 mV. The pH and ion concentrations exceed discharge regulations as promulgated by the U.S. Environmental Protection Agency (EPA) Gold Book standards for pristine water supplies by large factors, often several orders of magnitude. Table VII illustrates that this environmentally friendly solid supported macrocyclic system developed by IBC Advanced Technologies, Inc., of American Fork, Utah, was capable of achieving high metal recovery and effluent water qualities that meet EPA Gold Book and/or regulatory drinking water standards. The target metals were recovered without the addition of any undesirable ion to the system and were of sufficient purity

TABLE VII Silica Gel-Supported Macrocyclic Separation System Results from Superfund Site

Metal	Influent (Source) concentration (mg/L)	Treated effluent concentration[a] (mg/L)	Recovery (%)	Eluent purity (%)
Cu	180	<0.02	>99	96–98
Fe	994	<0.1	>99	99.5
Al	270	<1	>99[b]	97–98
Zn	554	<0.05	>99	99–99.5
Mn	194	<1	>99	75–80[b]
Cd	2	<0.02		
As	0.3	<0.1		

[a] All below detection by the analytical methods used.
[b] Some Al slipped past the Al system but was recovered with the Mn. Total Al recovery was >99%.

to allow them to be marketed as concentrated solutions for further refining.

There will be increased emphasis in the future on developing macrocyclic ligands that are highly selective for particular target cations and anions in many matrices and conditions. The impetus for these developments will be increased environmental awareness and the need for greater cost effectiveness in the separation and purification procedures.

C. Supramacrocycles: Self-Assembly and Molecular Machines

Tremendous advances in synthetic methodologies and macrocyclic structure exegesis have allowed the creation and characterization of macrocycles and surpramacrocycles which, until recently, were sequestered in the minds of creative chemists. Examples of molecular topology, intertwining, and self-assembly are ubiquitous in the natural world. For example, the double helix, lipids, viral capsids, and the secondary, tertiary, and quaternary structures of many biomolecules all posses varied levels of molecular topology, intertwining, and self-assembly. Macrocyclic chemistry of the 1970s has fathered the supramolecular chemistry of the 1980s and 1990s. It is now a new millennium and it is becoming feasible to construct large (nanometer to millimeter dimensions) and intricate supramolecular entities with machine-like properties in order to investigate the processes central to nature's forms and functions. Rotaxanes, pseudorotaxanes, and catenanes (Fig. 9) are interesting examples of this vast supramolecular class because the relative positions of their component parts can be induced to change as a result of an external stimulus. In most cases this mechanical mobility occurs between well-defined states that can be switched on or off by external stimuli such as chemical, photochemical, electrochemical, or electrical energy.

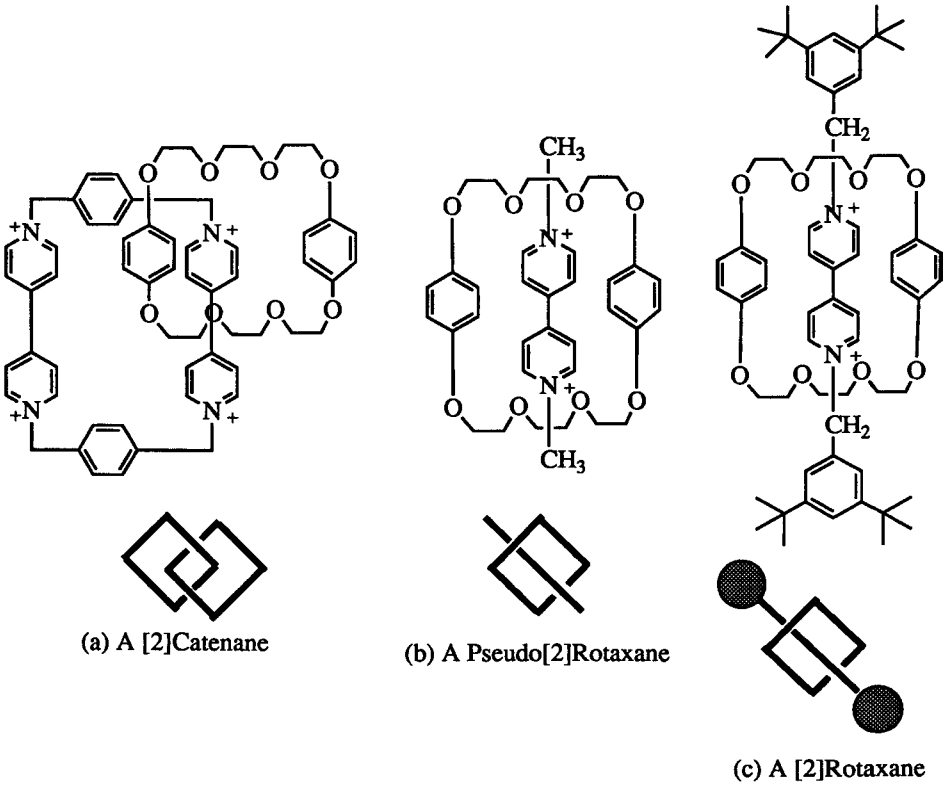


FIGURE 9 Examples of supramacrocycles.

Rotaxanes derive their name from the Latin word *rota*, meaning wheel, and *axis*, meaning axle. Rotaxanes consist of a dumbbell-shaped moiety, in the form of a rod and two bulky stopper groups around which there are encircling macrocyclic component(s). The sterically impeding stoppers of the dumbbell prevent the macrocycle(s) from disassociating from the rod portion of the assemblage. If these stopper groups are absent from the ends of the rod molecule, or of insufficient size to provide a steric barrier, the supramolecules are termed a pseudorotaxane.

Catenanes, from the Latin word *catena*, meaning chain, are molecules containing two or more interlocked rings that are inseparable unless a covalent bond is broken. The nomenclature used for these systems involves placing the number of components involved in square brackets prior to the name of the molecule. Thus, compound (a) in Fig. 9 is a [2]catenane and compound (c) in Fig. 9 is a [2]rotaxane.

Clearly, the first steps have been taken toward creating simple molecular machines. Figure 10 shows a molecular shuttle made from a [2]rotaxane capable of translational isomerism controlled by electrochemical and chemical switching. The dumbbell-shaped component in this example incorporates two different recognition sites, one a biphenol and the other a benzidine unit. The cyclophane ring system with its four formal positive charges preferentially locates at the more π-electron-rich benzidine unit in CD_3CN at $-44^{\circ}C$ (84:16) (middle of Fig. 10). Electrochemical oxidation of the benzidine unit converts it to a radical monocation state, and the tetracationic macrocyclic ring moves to the biphenol site (top of Fig. 10). This redox situation is completely reversible. Chemical switching also occurs by protonation (trifluoroacetic acid, TFA) to form a diprotonated benzidine species causing the tetracationic cyclophane ring to move to the biphenol site (bottom of Fig. 10). This chemical process is reversed upon addition of pyridine to remove the protons from the benzidine. Thus, a controllable molecular shuttle system has been developed. It must be emphasized that this is but one example in an incredibly large and extremely interesting field of supramolecular chemistry. Future research will surely focus on arrays and higher assemblies of molecules in this area.

V. SYNTHESIS OF MACROCYCLIC COMPOUNDS

The synthetic organic macrocyclic ligands come in all shapes and varieties. Early work was with the polynitrogen macrocyclic compounds such as the cyclams (compound 7, Fig. 1). More recent innovations have been with the macrocyclic polyether (crown) compounds. Charles J. Pedersen of duPont first reported these compounds in 1967. Pedersen was preparing the *bis*-phenol substituted polyether shown in Eq. (1). He isolated a good

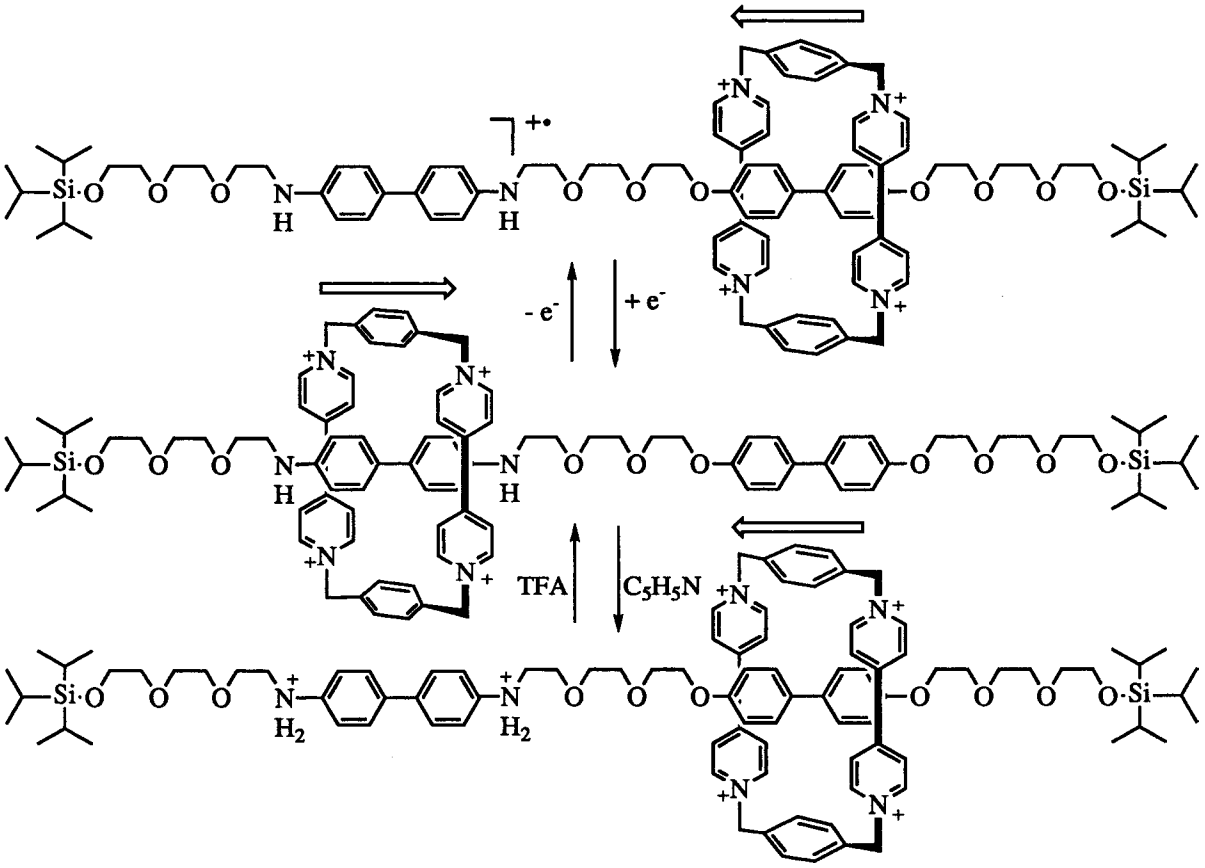
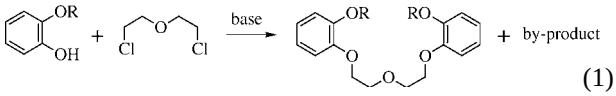
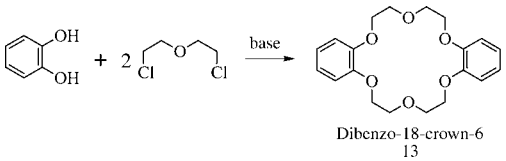


FIGURE 10 A molecular shuttle controllable by chemical and electrochemical switching.

yield of his intended product but persisted in purifying the by-product to obtain dibenzo-18-crown-6 (compound 13, Fig. 1) which proved to be a remarkable complexing agent for cations. Compound 13 was produced from the catechol impurity in the starting phenol shown in Eq. (1). When the same reaction was carried out with catechol, a good yield of compound 13 was isolated [Eq. (2)]. This reaction is a Williamsen ether synthesis.

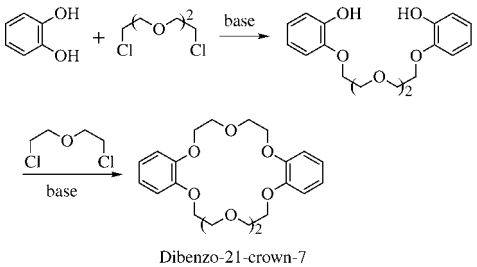


(1)

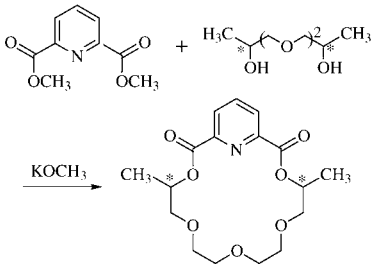


(2)

The synthesis of the crown compounds has been accomplished by a number of different cycloaddition methods. The basic 2 unit plus 2 unit addition as shown in Eq. (2) has been most used for the simple crowns. Often part of the macrocycle is first synthesized and then a second part added on, as in Eq. (3). This is particularly useful when preparing unsymmetrical crowns such as dibenzo-21-crown-7. The last part of the synthesis shown in Eq. (3) is a simple 1 unit to 1 unit cycloaddition. This process has been used to prepare other types of macrocyclic compounds such as pyridino diester-18-crown-6 (compound 12, Fig. 1) [Eq. (4)]. This is a transesterification reaction and gives excellent yields. Indeed, most of these cycloaddition reactions exhibit a template effect in that greater yields are realized when a cation, or in some cases, a neutral molecules, that fits into the macrocyclic cavity is used in the reaction.



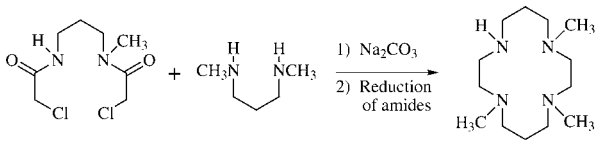
(3)



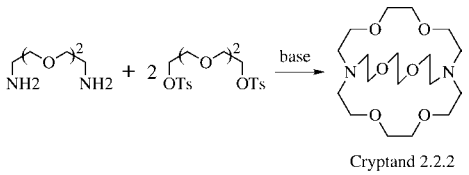
12 (4)

Compound **12** has shown chiral recognition for the enantiomers of various organic ammonium salts in that enantiomerically pure **12** forms a stronger complex with one enantiomer of a chiral organic ammonium salt than with the other enantiomer. Chiral crown 15 (Fig. 1) synthesized in a similar manner has been used to effect an NADH type of asymmetric reduction of certain aromatic ketones.

There are a great number of ways to prepare the azacrown ethers which include the cyclic polyamines such as cyclam (compound 7, Fig. 1). Some of these syntheses include the cycloaddition reaction of a diacid dichloride and a diamine followed by reduction of the macrocyclic diamide and treatment of a dihalide with a bistosylamide followed by removal of the *N*-tosyl groups. A more recent innovation to prepare a cyclam is the treatment of a diamine with a bis(α-chloroamide) as shown in Eq. (5) for the preparation of trimethylcyclam. The bis(α-chloroamide) is readily prepared from a diamine and chloroacetyl choride or chloroacetic anhydride. The two amide nitrogen atoms can contain hydrogens since the amide nitrogen cannot act as the nucleophile. Cryptand 2.2.2 (compound 4, Fig. 1) was initially prepared by Lehn and co-workers in a multistep process. A recent advance in the synthesis of macrobicyclic compounds is the one-step reaction of the appropriate bisprimary amine with two moles of a ditosylate as shown in Eq. (6). The ditosylate is the key to this one-step synthesis which results in the preparation of the cryptands in relatively high yields.

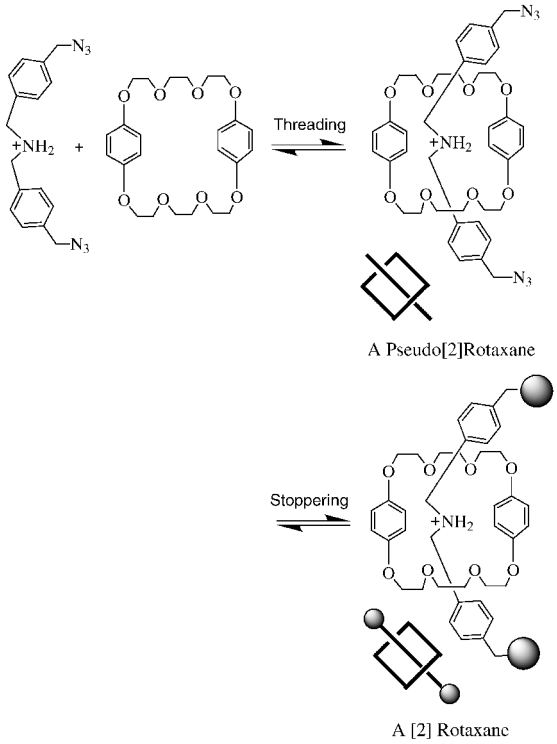


7 (5)



Cryptand 2.2.2
4 (6)

Synthetic supramolecular chemistry is currently receiving considerable attention. The creation of multicomponent macrocyclic architectures utilizing noncovalent bonding interactions and the synthesis of discrete molecular entities interlocked or intertwined by covalent bonds and mechanical associations assisted by intermolecular, noncovalent interactions is driven by the desire to understand and engineer well-defined, self-assembled structures with desirable properties. In the beginning, these supramolecular compounds were prepared in low yield via "statistical" methods. Advances in synthetic methodologies that take advantage of complexation and noncovalent association properties make the synthesis of these molecules more available. Self-assembly as a synthetic tool shows promise for supramacrocyclic compounds. A single representation of supramolecular synthesis is given in Eq. (7). The first step is the association of the organic ammonium cation with the crown ether in a process called threading. Then large bulky groups are linked to each end which prevents dethreading of the organic ammonium ion.



A Pseudo[2]Rotaxane

A [2] Rotaxane (7)

Thirty years after Pedersen's discovery of crown ethers, the host–guest concept has firmly established itself in the realm of synthetic chemistry, as evidenced by the expansion of the macrocyclic into the self-assembled supramolecular synthesis.

SEE ALSO THE FOLLOWING ARTICLES

INCLUSION (CLATHRATE) COMPOUNDS • LIGAND FIELD CONCEPT • MEMBRANES, SYNTHETIC (CHEMISTRY)

BIBLIOGRAPHY

Amabilino, D. B., and Stoddart, J. F. (1995). *Chem. Rev.* **95,** 2725–2828.

Balzani, V., Gomez-Lopez, M., and Stoddart, J. F. (1998). *Acc. Chem. Res*. **31,** 405–414.

Bradshaw, J. S., Krakowiak, K. E., and Izatt, R. M. (1993). "Aza-Crown Macorcycles," Wiley, New York.

Bradshaw, J. S., and Stott, P. E. (1980). *Tetrahedron* **36,** 461–510.

Dietrick, B. (1996). "Cryptands," in "Comprehensive Supramolecular Chemistry," Vol. 1 (Gokel, G. W., ed.), Pergamon, Oxford, New York, Tokyo.

Izatt, R. M., Pawlak, K., Bradshaw, J. S., and Bruening, R. L. (1995). *Chem. Rev*. **95,** 2529–2586.

Lehn, J.-M., Atwood, J. L., Davies, J. E. D., MacNicol, D. D., and Vögtle, F. (eds.) (1996). "Comprehensive Supramolecular Chemistry," Volumes 1–10, Pergamon, Oxford, New York, Tokyo.

Lindoy, L. F. (1989). "The Chemistry of Macrocyclic Ligand Complexes," Cambridge Univ. Press, Cambridge.

Reinhoudt, D. N. (ed.) (1999). "Supramolecular Materials and Technology," Wiley, New York, Seoul, Tokyo.

Zhang, X. X., Bradshaw, J. S., and Izatt, R. M. (1997). *Chem. Rev.* **97,** 3313–3361.

Organometallic Chemistry

Robert H. Crabtree
Yale University

GLOSSARY

Back donation Donation by the metal of one or more of its electron pairs into the ligand. Since an empty ligand orbital is required to allow this to happen, the ligand must usually be unsaturated, that is, have double or triple bonds.

Main group metals Metals from groups 1, 2, and 11–15 of Mendeleev's periodic table, such as Li, Mg, Al, Pb.

Organometallic compound A compound containing a metal-carbon bond.

Transition metals Metals from groups 3–10 of Mendeleev's periodic table, such as Ti, Fe, W, Pt.

ORGANOMETALLIC CHEMISTRY is the study of substances that contain an organic compound or fragment bound to a metal atom or ion by a metal-carbon (M–C) bond. By tradition, compounds containing a metal-hydrogen (M–H) bond are also included in organometallic chemistry. In certain non-English-speaking nations, such as Russia, the scope of the subject is extended beyond the metals and is called "organo-element chemistry."

Organometallic compounds behave somewhat differently depending on the type of metal involved. The main group metals are groups 1, 2, and 11—12, and the heavier elements of 13–15, and the transition metals are groups 3–10. The transition metals can bind a wider range of ligands and have more types of reactions available to them than do the main group metals.

I. MAIN GROUP ORGANOMETALLIC COMPOUNDS

A. With Metal Carbon Bonds

We will look at the main group cases first, the most important organometallic derivatives of which are the alkyls, such as LiMe, MeMgBr, $PbEt_4$, and similar species (Me = methyl, CH_3; Et = ethyl, C_2H_5). They are usually prepared either by treating an alkyl halide with the metal, or by the reaction of a metal halide with an alkylating agent, often an organometallic compound such as LiMe.

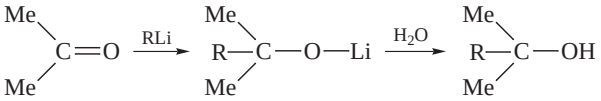
$$MeI + Li = LiMe + LiI \quad (1)$$

$$LiMe + CuCl = LiCuMe_2 + LiCl \quad (2)$$

The most electropositive metals, such as Na and K, tend to form ionic organometallic species, and therefore the best known are ones in which the anion of the organic fragment is unusually stable because the negative charge is delocalized over many atoms of the organic anion. For example, $Na[Ph_2CO]$ contains the $[Ph_2CO]^-$ radical anion; the unpaired electron carrying the negative charge is delocalized over the whole Ph_2CO molecule. There is no M—C bond per se, because the compound is ionic. These

species are good reducing agents; that is, they readily add an electron to a suitable acceptor molecule. They are very unstable to moisture and air.

Slightly less electropositive metals such as Li, Mg, Al do form covalent organometallic compounds. Many of these alkyls are very reactive, depending on the electronegativity of the metal involved, and are often rapidly decomposed by air and water. Aluminum compounds such as Al_2Me_6 burst into flames in air. The electropositive metals, such as Li, Mg, Zn, and Al, form alkyls that are exceptionally useful in synthetic organic chemistry because they introduce an R^- group into an organic compound, as shown in Scheme 1.

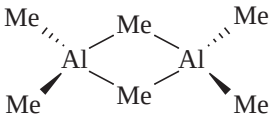


Scheme 1.

This reaction is strongly promoted by the binding of the metal to the oxygen end of the C=O bond, which polarizes it and so prepares this bond for R^- attack at the carbon.

The electronegative metals form much more stable and less reactive alkyls, for example, $HgMe_2$. These are also useful in organic chemistry, because the R group in an R_2Hg or RHgX (X = halide) compound can have any of a wide variety of functional groups, such as —COOMe, $—CONH_2$, organic carbonyl, —OH, etc. This is not true for RLi, because the C—Li bond reacts and is therefore incompatible with these groups.

Some of the group 1, 2, and 13 alkyls are dimeric or polymeric. The reason is that these elements have fewer than four electrons and so cannot reach the stable octet of electrons by forming alkyls MR_n, where n is the valency of the metal. Bridging allows the metal to reach the stable 8e configuration, as shown for Al_2Me_6 in Scheme 2. Main group organometallics tend to be tetrahedral, 4-coordinate, as shown for Al_2Me_6, but the heavier metals (At. No. > 10) can also attain higher coordination numbers.



Scheme 2.

Some of the later element organometallics have special applications. Organocopper compounds, for example, have a strong tendency to add in the unusual 1,4-fashion to α,β-unsaturated ketones, rather than in the usual 1,2-fashion (to give $CH_2{=}_2CMe(OH)R$) as would LiMe.

$$CH_2{=}CH{-}COR \xrightarrow{\text{i)LiCuMe}_2,\ \text{ii)H}_2\text{O}} MeCH_2CH_2COR \qquad (3)$$

The interest in new ways of making semiconductors, such as GaAs and InP, has led to the development of MOVD (metal-organic vapor deposition). In this application, two organometallic compounds are co-pyrolyzed in a ratio suitable for the deposition of the desired semiconductor. This low-temperature synthesis allows layers of semiconductors or insulators to be deposited in a controlled fashion and is being used for the fabrication of complex electronic components.

$$AsH_3 + GaMe_3 = GaAs + 4MeH \qquad (4)$$

Tetraethyl-lead, $PbEt_4$, was used in many countries as an antiknock agent in gasoline, but concerns about potential environmental problems led to its abandonment in favor of using a more highly branched chain and aromatics-rich gasoline.

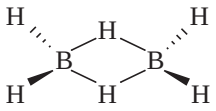
Methyl mercury cation, $HgMe^+$, is a water-soluble and unusually toxic form of this heavy metal. It can be formed from Hg^{2+} by bacterial action and was the causative agent in Minimata disease. Japanese families were poisoned because they were eating seafood caught in a bay where mercury wastes were being dumped. Mercury is normally present in the environment, although at much lower levels than those reached at Minimata. Certain bacteria have developed sophisticated enzymatic pathways to detoxify mercury compounds by reduction to metallic mercury, which is far less dangerous and will eventually evaporate from the cell. Higher animals, including humans, may also have proteins that protect from mercury and heavy-metal toxicity in general. The metallothioneins have been proposed to fill this role by binding heavy metals, but their function *in vivo* is still a matter of dispute.

B. With Metal Hydrogen Bonds

Main group hydrides are numerous. Once again, there is a gradation of properties across the periodic table, from the very reactive ionic compound KH to the modestly reactive, covalent SiH_4. The early hydrides, such as KH itself, are useful as nonnucleophilic bases, which only abstract a proton and do not add H^- to an organic compound. The later ones, especially in the anionic "-ate" complex form, are nucleophilic, for example, $NaBH_4$ and $LiAlH_4$. These contain the MH_4^- anion, which achieves its 8e configuration without bridging [M = 3e, H = 1e, anionic charge = 1e].

The neutral hydrides can have very interesting bridged structures. The boranes, B_xH_y, are the most extreme cases.

Hundreds of H-bridged species of this type are known, but the simplest is B_2H_6 (Scheme 3).



Scheme 3.

Diborane is one of the most useful main group organometallic species, because it adds to alkenes in the unusual anti-Markownikov direction to give organoboron derivatives that can be converted into any of a number of useful organic compounds.

$$RCH{=}CH_2 + 1/2B_2H_6 = RCH_2{-}CH_2BH_2 \quad (5)$$

$$RCH_2{-}CH_2BH_2 + H_2O_2 = RCH_2{-}CH_2OH \quad (6)$$

$$RCH_2{-}CH_2BH_2 + Br_2 = RCH_2{-}CH_2Br \quad (7)$$

Almost all other types of reagents that add to alkenes do so in the opposite direction to give the commercially less useful branched derivatives.

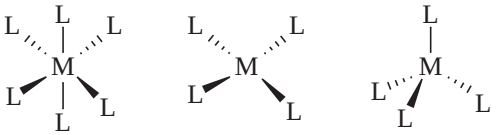
A development of interest in the area of main group organometallics was the isolation of compounds with multiple bonds between elements such as P and Si, e.g., RP═PR or $R_2Si{=}SiR_2$. Such heavy elements had been thought to be incapable of double bonding, and attempts to make compounds like these had always led to the formation of polymers such as $(PR)_n$. The solution turned out to be the use of very bulky R groups, which prevented polymerization. The resulting double bonds prove to have electronic structures interestingly different from those present in such long-known compounds as $R_2C{=}CR_2$.

II. TRANSITION METAL ORGANOMETALLIC COMPOUNDS

A. With Metal Carbon Bonds

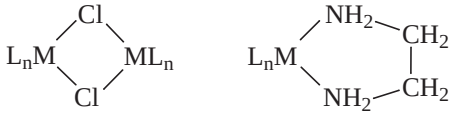
The transition metals, the elements from groups 3–10 of the periodic table, have been more intensively studied in recent years than the main group elements. Transition-metal organometallic chemistry grew out of coordination chemistry, of which it is a subfield. Coordination chemistry was founded in its modern form by Alfred Werner from 1896 and deals with compounds of metals and metal ions with ligands (symbolized in a general fashion as L). From the point of view of coordination chemistry, it is immaterial whether or not these ligands are bound by an M─C (or M─H) bond, and indeed the majority of coordination compounds have M─N or M─O bonds to the ligands. Metal ions are Lewis acids; that is, they can accept one or more pairs of electrons from one or more ligands L to form a complex or coordination compound $[ML_n]^{m+}$, such as $[Co(NH_3)_6]^{3+}$. The square brackets denote that the complex molecule (if $m = 0$) or ion (if $m \neq 0$) is an entity that retains its identity and can be regarded as a single molecule or ion. The M─L bonds are relatively strong. A key feature of coordination chemistry is that the presence of the ligands modify the properties of the central atom or ion, and the presence of the atom or ion modifies the properties of the ligand. The extent of these mutual modifications of metal and ligand can vary from minor to very profound. When we consider organometallic complexes, we find that this same mutual effect of metal on ligands and vice versa is also a very marked feature of the chemistry. One additional general property of complexes best illustrated by organometallic compounds is the possibility that ligands, usually of chemically different types, can react with one another within the coordination sphere of the metal (i.e., while still attached to the metal) in ways that are not observed for the free ligands L in the absence of a metal.

The commonest structural arrangements of the ligands around the metal in organometallic complexes in octahedral (**1**), square planar (**2**) or tetrahedral (**3**). Less regular arrangements can often be described as distortions of **1**–**3**.



1 **2** **3**

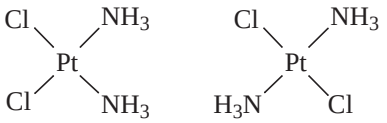
A ligand such as Cl^- or $H_2NCH_2CH_2NH_2$ has more than one lone pair of electrons (four on Cl^-, one on each of the two N atoms). In such a situation, two M─L bonds can be formed. Chloride ion commonly bridges two metals (**4**), while $H_2NCH_2CH_2NH_2$ more commonly chelates to a metal to form a ring (**5**) (L_nM refers to a metal and unspecified ligands). Chelate complexes are often more stable (thermodynamically and kinetically) than analogous species with monodentate ligands (i.e., ligands that have only one point of attachment to a metal).



4 **5**

Complexes that differ in the arrangement of the ligands around the central metal are different compounds and differ in properties. For example, *cis*-$[PtCl_2(NH_3)_2]$ (**6**) is an active anti-tumor drug, but *trans*-$[PtCl_2(NH_3)_2]$ (**7**) is not. Both **6** and **7** are square planar complexes, but they

are shown from the top in these pictures not from the side as in **2**.



6 **7**

The simplest carbon ligand is the methyl group, CH_3. Almost all metals form methyl complexes MMe_n, where n is the valency of the metal, which varies in a regular way across the periodic table (e.g., Na: $n = 1$. Mg: $n = 2$.). The M─CH_3 bond consists of a single shared pair of electrons and is not very different from the case of M─NH_3. Since carbon is more electronegative than the early metals (i.e., the ones on the left-hand side of the periodic table), the methyl group bears a negative charge in the methyl complexes of these elements (e.g., LiMe, mentioned earlier).

The next simplest carbon-based ligand is carbon monoxide, CO. Although superficially similar to the M─Me system in having a metal-carbon bond, M─CO is a very different type of bond. This molecule has a lone pair of electrons on the carbon atom, which can be shared with a metal to give a metal carbonyl complex, such as $Fe(CO)_5$. A new feature of the bonding, which differs from the situation in NH_3 or CH_3, is that the CO is an unsaturated compound (i.e., has double or triple bonds). In such a case, there is always the possibility of an additional interaction not present to any significant extent in complexes of NH_3 or the methyl group. The metal may donate some of its electron pairs to the ligand (an empty orbital allowing this to happen is almost always present in an unsaturated ligand); this is called back donation and accounts in part for the large modification both in metal and ligand properties on binding. For CO, the back bonding has the effect of weakening the C─O bond (this can be detected by infrared spectroscopy) and preventing the M─L bond from having the M^-─L^+ polarity of M─NH_3 (because for CO the ligand-to-metal electron donation is balanced by metal-to-ligand back bonding).

So important is this back donation in stabilizing the M─CO bond that metals which do not have any electrons to back donate form either no CO complexes or form very unstable ones. A metal will only bind CO strongly if it has d-electrons and therefore comes from groups 2–11 ($= N$) of the periodic table and has a valency state ($= v$), such that $N - v \geq 1$. Only unsaturated ligands (said to be soft ligands) have the power to accept back donation. Metals are able to engage in back bonding if they have a low valency state (e.g., $v \leq 2$). Soft metals include Pt(II), Mo(0), Fe(0), Ir(I), Hg(II), Cu(I), and Ag(I). Soft ligands include CO, ethylene, benzene, allyl, cyclopentadienyl.

A remarkable feature of the zerovalent metal carbonyls is that there is a regular change in their formulas and therefore in their constitution as we move from left to right across the periodic table: $Cr(CO)_6$, $(CO)_5Mn$─$Mn(CO)_5$, $Fe(CO)_5$, $(CO)_3Co(CO)_2Co(CO)_3$, $Ni(CO)_4$. Alternate compounds are mononuclear (i.e., contain one metal) but have one less CO for every two steps to the right. This regularity is embodied in the Eighteen Electron rule, which states that special stability often accompanies a valence electron count of 18 electrons per metal (and in some cases, such as Pd(II), 16 electrons). To take the case of $Cr(CO)_6$, the valence electron count is obtained as follows: Cr is in group 6 of the periodic table, so has 6e; CO donates the 2e of its C lone pair to the metal and so:

$$6 + (2 \times 6) = 18 \tag{8}$$

In the general case of a complex of formula $[MX_aL_b]^{c+}$ (where M is a metal of group N, X is an anionic ligand such as Cl or CH_3, and L is a neutral ligand, such as NH_3 or CO), the count is given by Eq. (9)

$$\text{e count} = N + a + 2b - c \tag{9}$$

Useful concepts are the oxidation state (O.S.), given by Eq. (10), and the d^n configuration, or number of d-electrons, given by Eq. (11).

$$\text{O.S.} = c + a \tag{10}$$

$$n = N - c + a \tag{11}$$

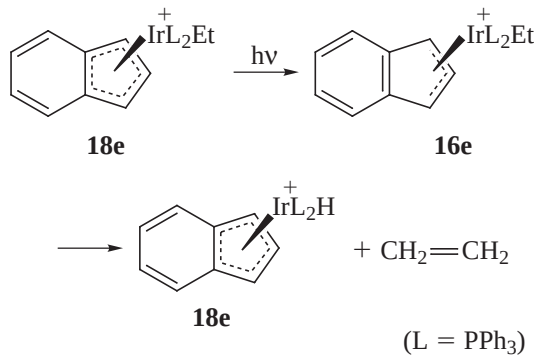
The (II) (i.e., $+2$) of Pd(II) is the oxidation state; Pd(II) is said to have the d^8 configuration.

In metal carbonyls, the CO removes metal electron density so efficiently that the lower oxidation states of the metal, which have more electrons, are strongly stabilized. Methyl complexes and carbonyls tend to have very different oxidation states for this reason, such as WMe_6 (O.S. = 6) and $W(CO)_6$(O.S. = 0).

Early metal alkyls tend to be reactive and nucleophilic (i.e., tend to transfer an alkyl anion to a reagent), but the late metals, such as Pt, Ir, Hg, can form very stable alkyls; the very high stability of the water-soluble $[HgCH_3]^+$ constitutes one of the major hazards of mercury pollution, because this species, formed by microbial action, can enter the food chain. Metal alkyls that have a β-H (i.e., a hydrogen at the second carbon from the metal) can undergo the β-elimination reaction provided they are coordinatively unsaturated (i.e., have an electron count of less than 18) or, if coordinatively saturated (18e count), can lose a ligand so as to generate the necessary unsaturation.

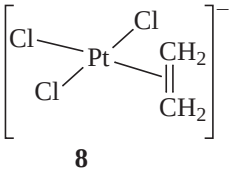
The reaction is much faster if the metal has a d^n configuration between 2 and 10, because d-electrons stabilize the transition state (or most unstable species along the reaction pathway; the more stable this is, the faster the reaction).

[(Indenyl)$L_2IrCH_2CH_3$] is an 18e d^6 species, but the indenyl can slip sideways on illumination so as to disengage 2e and generate unsaturation. As expected on the principles discussed above, illumination of [(Indenyl)$L_2IrCH_2CH_3$] with ultraviolet light leads to β-elimination, in which the β-H is transferred to the metal and (in this case) $CH_2{=}CH_2$ is released (Scheme 4).



Scheme 4.

We next examine unsaturated ligands, such as ethylene, C_2H_4. These are very soft ligands and only bind well to d^2–d^{10} metals. Ziese made the first one in 1837, but the structure (**8**) was only established in the 1950s.

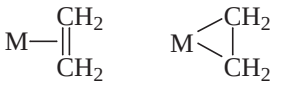


8

The ethylene acts as a neutral, 2e donor via its C=C π-bonding electrons, back donation (into the π^* levels) stabilizes the complex. Two extreme situations can be distinguished. If the back bonding is very restricted in its extent, the ethylene acts essentially only as a donor, and the ligand becomes depleted of electrons (**9a**, the Chatt–Dewar model). In such a situation the ligand can be attacked at the carbon atom by nucleophiles [e.g., Eq. (3)], a reaction that does not occur in the case of free ethylene. This is an example of the metal exerting a modifying effect on the reactivity of the ligand. This type of effect makes organometallic chemistry very useful in organic synthesis, the art of constructing organic molecules. A metal complex can be erected like a temporary scaffold around an organic compound so as to carry out a reaction that would not otherwise take place.

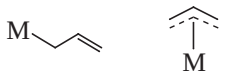
$$LCl_2Pt(C_2H_4) + Me_2NH = [LCl_2PtC_2H_4NMe_2]^- + H^+ \qquad (12)$$

In the opposite bonding extreme, back bonding is very important and the resulting complex can be considered as having two M–C single bonds, as shown in the metalacyclopropane model **9b**. Alkynes such as acetylene (HCCH) bind to metals in a similar way.



9a **9b**

The allyl group is an interesting ligand because it can either bind as a monodentate ligand (**10a**), rather like the methyl group, or it can bind via this bond and via the immediately adjacent C=C group as well (**10b**). If it binds in the first way it is denoted η^1-allyl; if in the second, η^3-allyl. In each case, the superscript indicates the number of ligand atoms bound to the metal, otherwise known as the hapticity. Variable hapticity is shown by a large number of organometallic ligands (e.g., Scheme 4).



10a **10b**

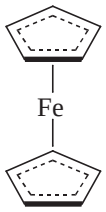
1,3-Dienes normally bind in an η^4-form, **11a**; both C=C bonds are effectively bound to the metal. When this happens an electronic redistribution takes place in the diene, so that the C_2–C_3 bond of the complexes diene becomes shorter than the C_1–C_2 and C_3–C_4 bonds, as illustrated by the resonance form **11b**; the C_2–C_3 bond is longer than the C_1–C_2 and C_3–C_4 bonds in the free ligand. This is another example of metal-induced change in a ligand. In this case the reason is that back bonding populates a molecular orbital, which leads to greater C_2–C_3 bonding in the complex than in the free diene.

11a **11b**

Cyclopentadienyl, C_5H_5 (=Cp) is one of the most celebrated ligands in organometallic chemistry. Unlike the unsaturated ligands mentioned up to now, it does not require back donation to bind strongly, and it forms a wide variety of complexes spanning the whole periodic table. η^1- (**12a**),

η^3- (**12b**), and η^5- (**12c**) structures are found, but the latter are by far the most numerous, as in ferrocene, **12d**.

12a **12b** **12c**



12d

The great advantage of this ligand is that it is unaffected by the usual reagents and so can be used as a spectator ligand to stabilize a metal fragment while other chemistry takes place. For example, the fragment $CpFe(CO)_2$, (=Fp) binds alkyl groups very efficiently, and it is used for this purpose in organic synthesis.

Arenes such as benzene normally bind in an η^6-mode, **13**, but only to soft metals; even then the arene dissociates easily. This difference with respect to Cp might at first appear puzzling, but the reason is that Cp is a poor leaving group; because Cp^- or $Cp^\bullet$ are relatively unstable, an arene can dissociate as the free arene, a very stable entity.

13

The preparation and characterization of organometallic compounds uses methods similar to those employed in organic chemistry. Notable in organometallic chemistry is the more common use of X-ray crystallography. Also notable are the strong spectral features due to metal carbonyls in the infrared spectrum and metal hydrides in the proton nuclear-magnetic-resonance (NMR) spectrum.

B. With Metal Hydrogen Bonds

Metal hydrides play an important role in organometallic chemistry. They add to a variety of unsaturated organic compounds by a reaction that is the reverse of the β-elimination step of Eq. (6) to give species with M–C bonds. Hieber made the first example of a transition-metal hydride in 1931, but his claim was not generally accepted until around 1960. Hydrides are now very easy to detect, thanks to the appearance of a characteristic peak in the nuclear magnetic resonance spectrum. Hydrides can be terminal, as in $(CO)_5Mn–H$, or bridging, as in $(CO)_5Cr–H–Cr(CO)_5$. H_2 can also be bound as a neutral 2e ligand to a metal, as in $[(CO)_5Mo–(H_2)]$, in which the H–H bond of the hydrogen molecule is retained.

C. Typical Reaction Types

The simplest reaction of an organometallic compound is a substitution in which one ligand replaces another at a metal center. A very common example is the replacement of a CO by a phospine, PMe_3 or PPh_3. Phosphines are very useful ligands because their bulk and electronic properties can be varied over a wide range in a controlled way; this allows us to explore how these affect the types of reactions we can bring about. In an 18e complex, such as $Mo(CO)_6$, a CO usually has to dissociate to give the 16e reactive intermediate $\{Mo(CO)_5\}$ before a new ligand can bind, as shown in Eq. (13), a dissociative substitution.

$$\begin{aligned} Mo(CO)_6 + PMe_3 &= \{Mo(CO)_5\} + PMe_3 + CO \\ &= [Mo(CO)_5(PMe_3)] + CO \end{aligned} \quad (13)$$

On the other hand, a 16e complex will usually give an associative substitution, in which the incoming ligand attacks to give an 18e intermediate, as shown in Eq. (14).

$$\begin{aligned} IrCl(PMe_3)_3 + CO &= \{IrCl(CO)(PMe_3)_3\} \\ &= [IrCl(CO)(PMe_3)_3] + PMe_3 \end{aligned} \quad (14)$$

Illumination sometimes causes a ligand to dissociate, and under these circumstances a photochemical substitution is often observed; carbonyls are especially prone to give this reaction.

A variety of compounds with reactive X–Y bonds can give the oxidative-addition reaction with any of a large number of reduced-metal species, as illustrated in Eqs. (15–17).

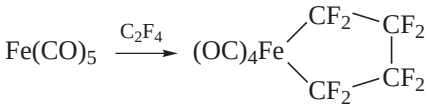
$$PPh_3 + Cl_2 = Cl_2PPh_3 \quad (15)$$

$$IrCl(PPh_3)_4 + H_2 = IrH_2Cl(PPh_3)_4 \quad (16)$$

$$Pd(PCy_3)_2 + MeI = MePdI(PCy_3)_2 \quad (17)$$

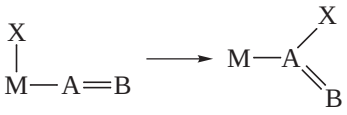
This reaction is a very versatile way of making M–L bonds. In spite of the similarity of the overall transformations, oxidative addition is a mechanistically diverse reaction class. In the reverse process, reductive elimination, an X–Y molecule is formed from a metal complex $L_nM(X)(Y)$.

A related process, oxidative coupling, is illustrated in Scheme 5.

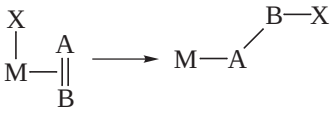


Scheme 5.

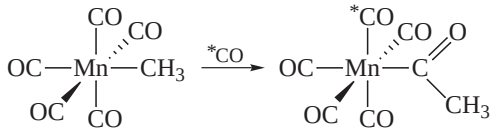
Insertion and elimination are mutually inverse reactions, in which a ligand, shown as AB, inserts into an M–X bond, illustrated in Schemes 6 and 7. Ligands that bind end-on usually give 1,1-insertions in which both M and X are attached to the same atom in the product. Ligands that bind side-on usually give 1,2-insertions. Some specific examples are shown: Scheme 8 is a 1,1 insertion to give an acetyl complex; Scheme 9 is a 1,2-insertion to give a vinyl complex; and Scheme 10 is a 1,2-insertion to give a sulfinate complex. β-Elimination, mentioned earlier, can now be seen to be a 1,2-deinsertion. It is notable that although β-elimination of an H is very rapid, there are very few examples of β-elimination of an alkyl group.



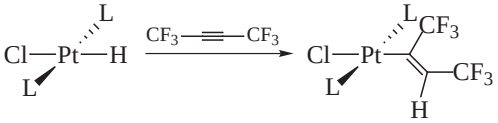
Scheme 6.



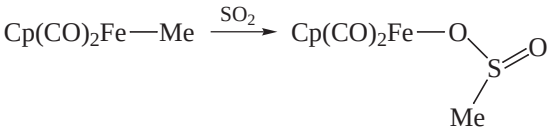
Scheme 7.



Scheme 8.

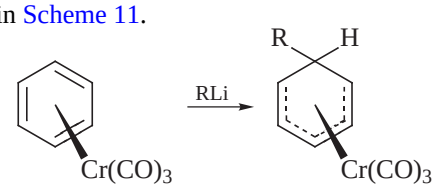


Scheme 9.



Scheme 10.

Nucleophilic addition to alkenes has been mentioned (Section I.B), but similar additions occur for other organometallic ligands. Especially facile are reactions with allyl, diene, and arene groups. For example, the arene complex $(C_6H_6)Cr(CO)_3$ reacts with MeLi as shown in Scheme 11.



Scheme 11.

An alternative way that a nucleophile can react is to abstract an X^+ fragment, as illustrated for H^+ abstraction in Eq. (18).

$$Cp_2TaMe_2^+ + R_3P{=}CH_2 = Cp_2Ta(=CH_2)Me + R_3P{-}CH_3^+ \quad (18)$$

An electrophile, in contrast, tends to add to uncoordinated parts of an unsaturated ligand (electrophilic addition) or abstract an X^- fragment (electrophilic abstraction). Equation (19) shows protonation of an η^1-allyl to give an η^2-alkene cation, and Eq. (20) shows the protonation of a hydroxyethyl complex.

$$Fp{-}CH_2{-}CH_2{=}CH_2 + H^+ = [Fp(CH_2{=}CH_2{-}CH_3)]^+ \quad (19)$$

$$Fp{-}CH_2{-}CH_2{-}OH + H^+ = [Fp(CH_2{=}CH_2)]^+ + H_2O \quad (20)$$

Nucleophilic addition to coordinated CO is an important way of making carbenes, as we will see later.

D. Applications in Catalysis

Homogeneous catalysis is an important commercial application of organometallic chemistry; many millions of pounds of a wide variety of organic compounds are made in this way each year. The object is to find an organometallic complex which, when present in small, even trace, amounts, will bring about the conversion of a given reagent into a desired product. To do this, the metal has to coordinate to the substrate (i.e., organic reagent) or substrates and activate them for the desired reaction. In this area of chemistry, the intention is to mimic certain of the properties of enzymes, but in contrast to many enzymes, organometallic catalysts tend to be robust and accept a wide range of substrates. Catalysis is an example of "green," or environmentally conscious, chemistry, because it reduces the amount of waste products formed.

Alkene hydrogenation is an area in which several different catalysts have been shown to be useful for different purposes. For example, $RhCl(PPh_3)_3$ is selective for unhindered C=C double bonds and causes very little isomerization in the substrate. $[Ir(cod)(PCy_3)py]^+$, on the other hand, is very active for very hindered C=C groups, and if

a binding group such as –OH is present in the alkene, the catalyst will bind to this group and then add the H_2 to the C=C bond. This alters the stereochemistry of the product in a useful way. A simplified mechanistic scheme for the reaction is shown for ethylene and H_2 in Eqs. (21–24). An oxidative addition of H_2 leads to the dihydride. The ethylene then binds and inserts into one M–H bond to give the metal ethyl. The ethyl group then reductively eliminates with the remaining H ligand to give ethane. The final reductive elimination is the reverse of the initial oxidative addition, but involves a C–H rather than an H–H bond. Note that the catalyst, symbolized by L_nM, is regenerated in the final step so that it can react with a second and subsequent molecules of the reagents. Many thousands or millions of catalytic turnovers can be carried out, which means that very little catalyst needs to be present.

$$L_nM + H_2 = L_nMH_2 \quad (21)$$

$$L_nMH_2 + H_2C{=}CH_2 = L_nMH_2(H_2C{=}CH_2) \quad (22)$$

$$L_nMH_2(H_2C{=}CH_2) = L_nMH(CH_2{-}CH_3) \quad (23)$$

$$L_nMH(CH_2{-}CH_3) = L_nM + H_3C{-}CH_3 \quad (24)$$

Hydroformylation, Eqs. (25) and (26), is useful for the preparation of aldehydes and, by hydrogenation of the aldehydes *in situ*, alcohols.

$$RCH{=}CH_2 + CO + H_2 = RCH_2{-}CH_2{-}CHO \quad (25)$$

$$RCH_2{-}CH_2{-}CHO + H_2 = RCH_2{-}CH_2{-}CH_2OH \quad (26)$$

The mechanism is similar to that for hydrogenation, but at the alkyl hydride stage a carbonyl insertion reaction takes place to lead to the sequence shown in Eqs. (27)–(29).

$$L_nMH(CH_2{-}CH_3) + CO = L_{(n-1)}M(CO)H(CH_2{-}CH_3) \quad (27)$$

$$L_{(n-1)}M(CO)H(CH_2{-}CH_3) = L_{(n-1)}MH(COCH_2{-}CH_3) \quad (28)$$

$$L_{(n-1)}MH(COCH_2{-}CH_3) = L_{(n-1)}M + HCOCH_2{-}CH_3) \quad (29)$$

Some hydroformylation catalysts are also alkene isomerization catalysts; that is to say they move the position of the C=C double bond along a chain, as shown in Eq. (30). This can happen much faster than hydroformylation itself. Advantage can be taken of this effect if one of the isomers reacts faster in the hydroformylation sequence than the others. In the case shown in Eqs. (30) and (31) the terminal alkene reacts faster and forms the linear aldehyde shown. Of three possible aldehydes, only one is formed in significant quantities if the right catalyst is used. This high reaction selectivity is a useful property of homogeneous catalysts.

$$CH_3CH{=}CHCH_3 = CH_3CH_2CH{=}CH_2 \quad (30)$$

$$CH_3CH_2CH{=}CH_2 + H_2 + CO = CH_3CH_2CH_2CH_2CHO \quad (31)$$

Alkene hydrocyanation, the addition of HCN across an alkene, is useful in the industrial preparation of adiponitrile [Eq. (32)], the key intermediate in the manufacture of nylon.

$$CH_2{=}CHCH{=}CH_2 + 2HCN = NC{-}CH_2CH_2CH_2CH_2CN \quad (32)$$

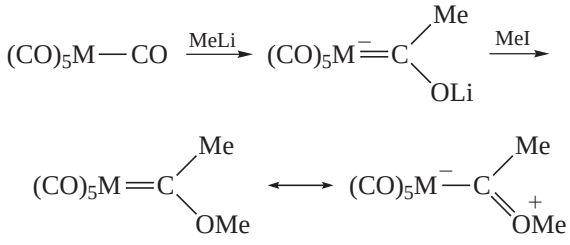
Hydrosilation, the addition of a silane, $R_3Si{-}H$ across an alkene is useful in the industrial preparation of silicone polymers and materials.

$$RCH{=}CH_2 + R_3Si{-}H = RCH_2CH_2SiR_3 \quad (33)$$

Polymerization of alkenes is an important area in which homogeneous catalysts, in the form of Ziegler–Natta and metallocene systems, have been very important. These catalysts are formed from an early metal compound, such as Cp_2ZrCl_2 or $TiCl_4$, an aluminum alkyl, such as Me_2AlCl, and a trace of water. They are believed to operate by successive insertion reactions, as shown in Eq. (34).

$$L_nM(CH_2CH_3) + H_2C{=}CH_2 = L_nM(CH_2CH_2CH_2CH_3) \quad (34)$$

Metal-carbon multiple bonds are found in a number of situations. The simplest is illustrated in Scheme 12, which shows the synthesis of a carbene complex. When a heteroatom is present, α to the carbene carbon, the M–C bond is not a full double bond because both resonance structures (Scheme 12) are thought to contribute. Carbenes without heteroatom substituents have been prepared, and these seem to have a true double bond. One example is $Cp_2Ta(=CH_2)Me$, formed by deprotonation of the $Cp_2TaMe_2^+$ cation.

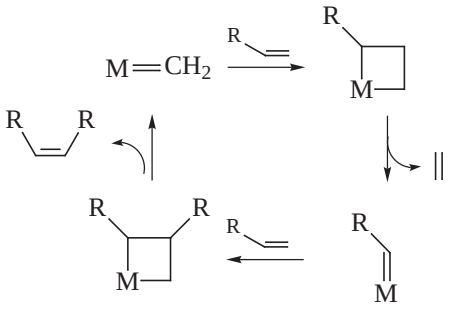


Scheme 12.

M=C double bonded intermediates are believed to be important in alkene metathesis Eq. (35), a useful commercial process that has been applied to organic synthesis, polymerization, and to changing the molecular weight distribution of a mixture of alkenes (Eq. 35).

$$RCH{=}CH_2 = RCH{=}CHR + CH_2{=}CH_2 \quad (35)$$

The key cyclic intermediate in this process is shown in Scheme 13.

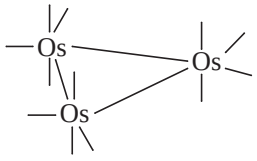


Scheme 13.

The Shell Higher Olefins Process (SHOP) is an example of a large-scale commercial process that relies on a metathesis step. In SHOP, ethylene is oligomerized to alkenes having a chain length of around C_{16} by a homogeneous nickel catalyst. The chain lengths of the resulting mixture can usefully be manipulated via metathesis.

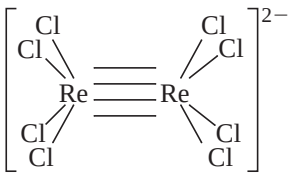
E. Compounds with Metal–Metal Bonds

Some compounds containing a metal–metal bond occur in organometallic chemistry. M—M bonds are important in "clusters," such as $Os_3(CO)_{12}$, which contains a triangle of metal atoms. The CO ligands are shown only as lines radiating from the triosmium core in the diagram. Clusters contain three or more metals with at least two M—M bonds (see Scheme 14).



Scheme 14.

The second situation in which M—M bonds are important is the case of dinuclear species. Multiple bonding is more often found in dinuclear cases, for example $Re_2Cl_8^{2-}$ (Scheme 15), but there are not as yet very many metal–metal multiply bonded organometallic compounds.

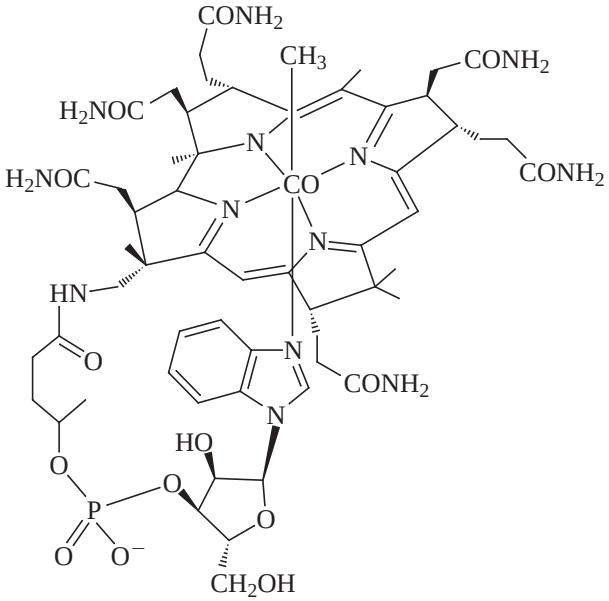


Scheme 15.

A number of useful electron counting rules, analogous to the 18e rule for mononuclear compounds, enables us to predict the structures to be expected among metal clusters. Nobelist Roald Hoffmann has developed a series of analogies between metal fragments and organic fragments that is another useful structural and conceptual tool. For example, $Os(CO)_4$ is said to be isolobal with CH_2, and so the triosmium cluster shown in Scheme 14 is isolobal with cyclopropane.

F. Organometallic Species in Biology

M—C bonds are proving to be important in biology. The best-known case is vitamin B-12, one derivative of which is shown in Scheme 16. The form shown is one of several derivatives that contains a metal–carbon bond. In one other naturally occurring organometallic derivatives of B-12, an adenosyl group can replace the methyl group. Lack of B-12 leads to the fatal disease pernicious anemia. The reason is that coenzyme B-12 is a cofactor in a number of vital steps in human biochemistry, for example synthesis of deoxy-ribose for DNA, and the synthesis of the amino acid methionine. B-12 contains cobalt, and its biochemical role involves making and breaking of Co—C bonds. It has also been found that Ni—C and Ni—H bonds are important in bacterial biochemistry, where they allow certain bacteria to live on H_2 and CO as energy and carbon source and to produce methane. Methane, or "marsh gas," is a common natural product; almost all of it appears to be formed by a bio-organometallic route.



Scheme 16.

G. Recent Advances and Current Problems

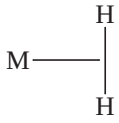
One of the areas of greatest current interest is the organometallic chemistry of alkanes. Alkanes are abundant and cheap, yet methods for their conversion into useful derivatives are lacking. For example, methane is flared off in the Sahara because of transport problems for this gaseous fuel. A way to combine it with air to give the easily transportable fuel methanol would be of value. It is hoped that organometallic catalysts will be found that can transform methane into methanol. Although certain bacteria can carry out this reaction, the chemist cannot yet do the same thing efficiently in the laboratory.

$$CH_4 + O_2 = CH_3OH \qquad (36)$$

It is not so much that CH_4 is unreactive, because it will easily burn in air to give CO_2. Rather it is the *intermediates* in the air oxidation that would be valuable if the oxidation could be stopped before the carbon dioxide stage. The reason this is exceptionally difficult is that these intermediates are much more reactive than methane itself.

Nature solves the problem by designing an enzyme that binds CH_4 strongly but rejects CH_3OH from the reaction site. This is done by arranging for the active site to be very nonpolar, a circumstance favorable to the binding of the nonpolar methane but not to binding of methanol, a polar molecule.

Another area of current interest is the attempt to prepare hydrides with H─H bonds in them. H_2 is the simplest molecule, and so its binding to metals is of unusual interest. The resulting structure is shown in Scheme 17. Scores of similar cases have now been discovered, and chemists are interested to see whether larger assemblies of hydrogen atoms can be stabilized by binding to a metal; for example, could an H_3 complex be made? One of the more challenging features of this search is that it is exceedingly difficult to distinguish between the well-known polyhydrides L_nMH_x, where there are M─H bonds only, and the interesting "nonclassical" form containing an H_x ligand. This is because the properties of the two might be very similar, and the method usually used to determine structures, X-ray crystallography, is very poor at determining H positions, because X-rays interact poorly with light atoms like H. The much more arduous and difficult neutron-diffraction experiment can locate hydrogen atoms precisely, because neutrons interact relatively strongly with all nuclei, whether of light or of heavy atoms. This technique can only be used sparingly, however, because there are only a very few neutron-diffraction facilities around the world, and so suitable candidate molecules must be located by other means.



Scheme 17.

SEE ALSO THE FOLLOWING ARTICLES

CATALYSIS, HOMOGENEOUS • MAIN GROUP ELEMENTS • METALORGANIC CHEMICAL VAPOR DEPOSITION • ORGANIC CHEMISTRY, SYNTHESIS • POLYMERS, INORGANIC AND ORGANOMETALLIC

BIBLIOGRAPHY

Albert, M. R., and Yates, J. T. (1987). "A Surface Scientist's Guide to Organometallic Chemistry," American Chemistry Society, Washington, D.C.

Crabtree, R. H. (1994). "Organometallic Chemistry of the Transition Elements," Wiley, New York 2nd ed.

Jenkins, P. R. (1992). "Organometallic Reagents in Organic Synthesis," Oxford, New York.

Schlosser, M. (1994). "Organometallics in Synthesis," Wiley, New York.

Wilkinson, G., ed. (1982–1994). "Comprehensive Organometallic Chemistry, I and II," Pergamon, Oxford.

Yamamoto, A. (1986). "Organometallic Chemistry: Fundamental Concepts and Applications" (Engl. Trans.), Wiley, New York.

Pharmaceuticals, Controlled Release of

Giancarlo Santus
Recordati Industria Chimica e Farmaceutica S.p.A.

Richard W. Baker
Membrane Technology and Research, Inc.

GLOSSARY

Biodegradable polymers Materials that undergo slow chemical degradation in the body, finally degrading to low-molecular-weight fragments that dissolve or are metabolized.

Controlled release systems Devices that meter delivery of drugs or other active agents to the body at a rate predetermined by the system design and are largely unaffected by the surrounding biological environment.

Enteric coatings pH-Sensitive materials used to delay dissolution of tablets in the acid environment of the stomach but allow rapid dissolution when the tablet reaches the more neutral pH environment of the gastrointestinal tract.

Microcapsules Small, drug-containing particles with diameters between 50 and 1000 μm.

Targeted drug delivery Delivery of a drug directly to the body site where the drug has its biological effect.

Transdermal patches Drug-containing laminates attached to the skin with an adhesive. Drug contained in the laminate migrates from the patch through the skin and is absorbed into the blood circulatory system.

Zero order delivery (of drug) Constant drug delivery over a certain period of time.

THE OBJECTIVE OF CONTROLLED drug delivery devices is to deliver a drug to the body at a rate predetermined by the design of the device and independent of the changing environment of the body. In conventional medications, only the total mass of drug delivered to a patient is controlled. In controlled drug delivery medications, both the mass of drug and the rate at which the drug is delivered is controlled. This additional level of control enhances the safety and efficacy of many drugs. Often a membrane is used to moderate the rate of delivery of drug. For example, in some devices, a membrane controls permeation of the drug from a reservoir to achieve the drug delivery rate required. Other devices use the osmotic pressure produced by diffusion of water across a membrane to power miniature pumps. In yet other devices, the drug is impregnated into a polymer material, which then slowly dissolves or

degrades in the body. Drug delivery is then controlled by a combination of diffusion and biodegradation.

I. INTRODUCTION/HISTORY

The pharmaceutical industry has produced long-acting oral medications since the 1950s. Enteric coated tablets were the first long-acting medication to be widely used. In 1952, Smith Kline French (SKF) introduced Spansules, or "tiny time pills," an over-the-counter cold medication consisting of millimeter-sized pills containing the drug and covered with a coating designed to delay its dissolution. By varying the thickness of the coating, different drug release profiles were achieved. These early products are best called *sustained release* systems, meaning that the release of the drug, although slower and more controlled than for a simple, fast-dissolving tablet, was still substantially affected by the external environment into which it was released. In contrast, the release of drug from *controlled release* systems is controlled by the design of the system and is largely independent of external environmental factors.

The founding of Alza Corporation by Alex Zaffaroni in the 1960s gave the development of controlled release technology a decisive thrust. Alza was dedicated to developing novel controlled release drug delivery systems. The products developed by Alza during the subsequent 25 years stimulated the entire pharmaceutical industry. The first pharmaceutical product in which the drug registration document specified both the total amount of drug in the device and the delivery rate was an Alza product, the Ocusert, launched in 1974. This device was designed to delivery the drug pilocarpine to control the eye disease glaucoma. The device consisted of a thin, elliptical, three-layer laminate with the drug sandwiched between two rate-controlling polymer membranes through which the drug slowly diffused. It was placed in the cul de sac of the eye, where it delivered the drug at a constant rate for 7 days, after which it was removed and replaced. The Ocusert was a technical tour de force, although only a limited marketing success. Alza later developed a number of more widely used products, including multilayer transdermal patches designed to deliver drugs through the skin. The drugs included scopolamine (for motion sickness), nitroglycerin (for angina), estradiol (for hormone replacement), and nicotine (for control of smoking addiction). Many others have followed Alza's success, and more than 20 transdermal products, delivering a variety of drugs, are now available. Alza also developed the first widely used osmotic controlled release drug delivery systems under the trade name Oros. The first billion-dollar controlled release product was an osmotic product, Procardia XL, delivering the drug nifedipine, a widely prescribed antihypertensive. Other important products launched in the last 15 years include Prilosec, a diffusion-controlled microparticle oral formulation system for the drug omeprazole, and Lupron (leuprolide) and Zoladex (goserelin), polypeptide drugs delivered from biodegradable intramuscular and subcutaneous implants. Since the first controlled release product was launched, the controlled release industry has grown to a $10–20 billion industry with more than 30 major controlled release products registered with the U.S. Food and Drug Administration. A time-line showing some of the important milestones in the growth of controlled release technology is shown in Fig. 1. Development of the technology has been reviewed in a number of monographs.

Controlled slow release of drugs to the body offers several important benefits to the patient.

- More constant drug blood levels. In controlled release products, drug delivery to the body is metered slowly over a long period, avoiding the problems of overdosing and underdosing associated with conventional medications. These problems are illustrated in Fig. 2, which shows the blood levels achieved when a relatively rapidly metabolized drug is given as a series of conventional fast-dissolving tablets. Immediately after the drug is ingested, blood levels begin to rise, reaching a peak value, after which the concentration falls as drug is metabolized. In Fig. 2 two important concentration levels are shown: the minimum effective concentration, below which the drug is ineffective and the toxic concentration, above which undesirable side effects occur. Maintaining the concentration of the drug between these two levels is critical for safety and effectiveness. Controlled release systems meter delivery of the drug to the body at a rate equal to the rate of drug removal by metabolization, so that a prolonged constant blood level is maintained. Because controlled release products use the drug more efficiently, the total amount required to produce the desired effect is often very much reduced, frequently by a factor of 10 or more.
- Improved patient compliance. A second benefit of controlled release formulations is improved patient compliance. Many studies have shown that few patients take their medications in complete accordance to physician instructions, and that the degree of noncompliance increases significantly as the complexity of the instructions increases. In one study of patients given once-per-day estrogen/progesterone contraceptive tablets in a controlled fashion, the pregnancy rate was less than 1 per 1000 patient-years. The same tablets prescribed to a normal population of women resulted in a pregnancy rate of 50 per 1000 patient-years due to noncompliance. Delivery of the similar contraceptive steroid levonorgestrel from a sustained release implant resulted in a pregnancy rate of

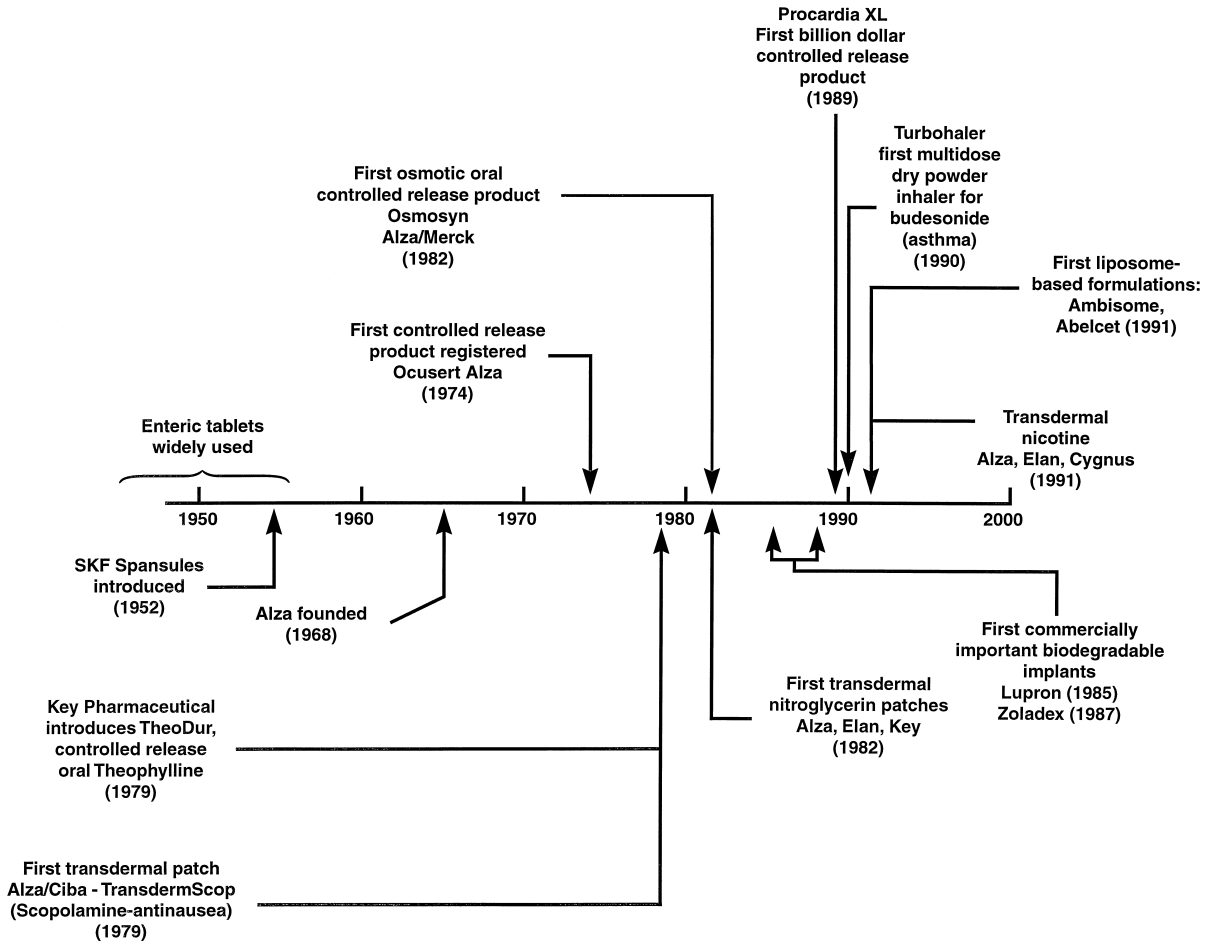


FIGURE 1 Milestones in the development of controlled release pharmaceuticals.

0.5 per 1000 patient-years. This product has the lowest pregnancy rate, that is, the highest degree of contraceptive efficiency, of any steroidal contraceptive because it eliminates poor patient compliance, the principal reason for drug failure.

- Targeted site of action delivery. A final benefit of controlled release formulation is the ability to deliver drug directly to the systemic circulation. Drugs taken orally are absorbed into the blood stream in the gastrointestinal (GI) tract. The drug then enters the portal circulation, so is first taken to the liver before entering the systemic circulation and being transported to the site of drug action. The liver's function is to deactivate toxic agents that enter the body through the GI tract. Unfortunately, the liver often treats drugs as toxic agents and metabolizes a large fraction before the drug reaches the blood circulatory system; this is called the first-pass effect. For example, in the first pass through the liver, 70–90% of the hormone estradiol is lost. Therefore, when used for hormone replacement therapy, it must be administered as tablets of 1–2 mg/day to achieve effective systemic blood levels. When the same drug is delivered directly to the systemic blood circulation from a transdermal patch, a delivery rate of only 50 μg of estradiol/day is sufficient to achieve the required effect.

II. METHODS OF ACHIEVING CONTROLLED RELEASE

A. Membrane Diffusion-Controlled Systems

In membrane diffusion-controlled systems, a drug is released from a device by permeation from its interior reservoir to the surrounding medium. The rate of diffusion of the drug through the membrane governs its rate of release. The reservoir device illustrated in Fig. 3 is the simplest type of diffusion-controlled system. An inert membrane encloses the drug, which diffuses through the membrane at a finite, controllable rate. If the concentration of the material in equilibrium with the inner surface of the enclosing membrane is constant, then the concentration gradient, that is, the driving force for diffusional release of the drug, is constant as well. This occurs when the inner

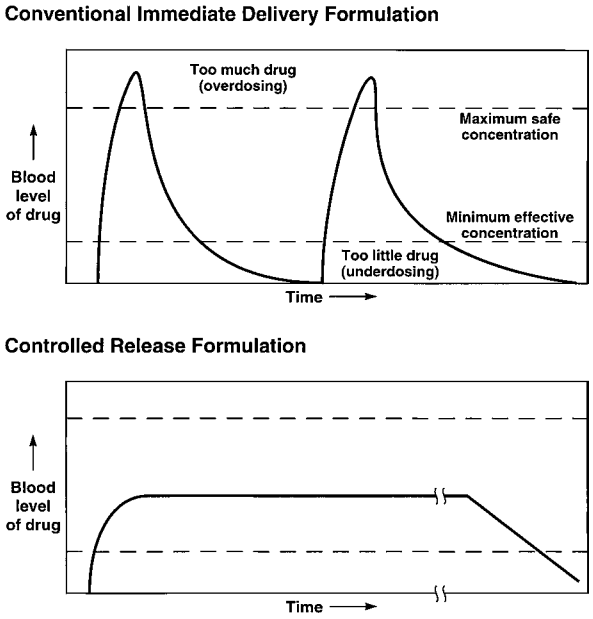


FIGURE 2 Simplified blood level profile illustrating the difference between repeated treatment with a conventional instant delivery formulation and a single, controlled release, long-acting formulation.

reservoir contains a saturated solution of the drug, providing a constant release rate for as long as excess solid is maintained in the solution. This is called zero-order release. If, however, the active drug within the device is initially present as an unsaturated solution, its concentration falls as it is released. The release rate then declines exponentially, producing a first-order release profile. The drug release profile for a simple membrane reservoir system is also shown in Fig. 3.

The drug release rate from a membrane reservoir device depends on the shape of the device, which may be a simple laminate, a cylindrical device, or a spherical device. For the simple laminate, or sandwich geometry, the drug release rate can be written

$$\frac{dM_t}{dt} = \frac{AP\Delta c}{l}, \tag{1}$$

where dM_t/dt is the device release rate, A is the total area of the laminate, P is the membrane permeability, l is the membrane thickness, and Δc is the difference in drug concentration between the solution at the inside surface of the membrane and the drug concentration in the solution at the outer surface of the membrane, usually close to zero. When the solution inside the enclosure is saturated, the drug concentration is c_s, and Eq. (1) reduces to

$$\frac{dM_t}{dt} = \frac{APc_s}{l}. \tag{2}$$

Drug release is then constant as long as a saturated solution is maintained within the enclosure. The total duration of constant release depends on the initial volume of the enclosure V, the mass of encapsulated drug M_0, and the solubility of the drug c_s. The mass of agent that can be delivered at a constant rate is $M_0 - c_sV$. Thus, it follows that the duration of constant release t_∞ is

$$t_\infty = \frac{M_0 - c_sV}{dM_t/dt}. \tag{3}$$

A second type of diffusion-controlled system is a monolithic or matrix device in which the drug is dispersed uniformly throughout the rate-controlling polymeric medium. The drug release rate is then determined by its loading in the matrix, the matrix material, the shape of the device (flat disk, cylinder, or sphere), and the permeability of drug in the matrix material. Equations describing release from all the various device types and geometries can be found elsewhere. As an example, desorption of drug uniformly dissolved in a simple disk (slab)-shaped device can be expressed by either of the two series

$$\frac{M_t}{M_0} = 4\left(\frac{Dt}{l^2}\right)^{1/2}\left[\pi^{-1/2} + 2\sum_{n=0}^{\infty}(-1)^n \,\text{ierfc}\left(\frac{nl}{2\sqrt{Dt}}\right)\right] \tag{4}$$

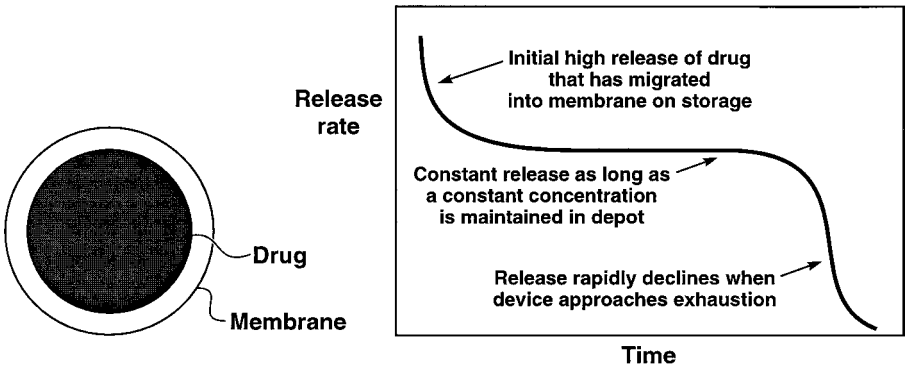


FIGURE 3 Schematic and drug release profile of the simplest form of membrane diffusion-controlled drug delivery system.

or

$$\frac{M_t}{M_0} = 1 - \sum_{n=0}^{\infty} \frac{8\exp[-D(2n+1)^2\pi^2 t/l^2]}{(2n+1)^2\pi^2}, \quad (5)$$

where M_0 is the total amount of drug sorbed, M_t is the amount desorbed at time t, and l is the thickness of the device.

Fortunately, these expressions reduce to two much simpler approximations, reliable to better than 1%, valid for different parts of the desorption curve. The early-time approximation, which holds over the initial portion of the curve, is derived from Eq. (4):

$$\frac{M_t}{M_0} = 4\left(\frac{Dt}{\pi l^2}\right)^{1/2} \quad \text{for} \quad 0 \le \frac{M_t}{M_0} \le 0.6. \quad (6)$$

The late-time approximation, which holds over the final portion of the desorption curve, is derived from Eq. (5):

$$\frac{M_1}{M_0} = 1 - \frac{8}{\pi^2}\exp\left(\frac{-\pi^2 Dt}{l^2}\right) \quad \text{for} \quad 0.4 \le \frac{M_t}{M_0} \le 1.0. \quad (7)$$

The release rate is easily obtained by differentiating Eqs. (6) and (7) to give

$$\frac{dM_t}{dt} = 2M_0\left(\frac{D}{\pi l^2 t}\right)^{1/2} \quad (8)$$

for the early time approximation and

$$\frac{dM_t}{dt} = \frac{8DM_0}{l^2}\exp\left(-\frac{\pi^2 Dt}{l^2}\right) \quad (9)$$

for the late time approximation.

These two approximations are plotted against time in Fig. 4. For simplicity, M_0 and D/l^2 have been set to unity. The release rate falls off in proportion to $t^{-1/2}$ until 60% of the agent has been desorbed, after which the decay is exponential. Although the release rate from monolithic devices is far from constant, this defect is often offset by their ease of manufacture.

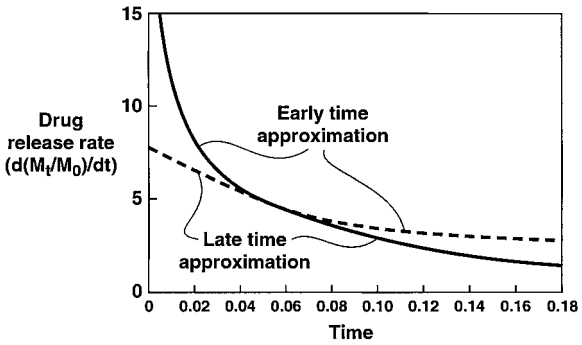


FIGURE 4 Drug release rates as a function of time, showing early- and late-time approximations.

B. Biodegradable Systems

The diffusion-controlled devices outlined above are permanent, in that the membrane or matrix of the device remains in place after its delivery role has been completed. In applications in which the controlled release device is an implant, the drug-depleted device shell must be removed after use. This is undesirable; such applications require a device that degrades during or after completion of its delivery role.

Several polymer-based devices that slowly biodegrade when implanted in the body have been developed. The most important materials are those based on polylactic acid, polyglycolic acid, and their copolymers, as shown in Fig. 5. Other, less widely used biodegradable materials include the poly(ortho esters), polycaprolactone, polyanhydrides, and polycarbonates.

In principle, the release of an active agent can be programmed by dispersing the material within such polymers, with erosion of the polymer effecting release of the agent. One class of biodegradable polymers is *surface eroding*: the surface area of such polymers decreases with time as the cylindrical- or spherical-shaped device erodes. This results in a decreasing release rate unless the geometry of the device is appropriately manipulated or the device is designed to contain a higher concentration of the agent in the interior than in the surface layers. In a more common class of biodegradable polymer, the initial period of degradation occurs slowly. Thereafter, the degradation rate increases rapidly and the bulk of the polymer then erodes in a comparatively short time. In the initial period of exposure to the body, the polymer chains are being cleaved but the molecular weight remains high. Therefore, the mechanical properties of the polymer are not seriously affected. As chain cleavage continues, a point is reached at which the polymer fragments become swollen or soluble in water; at this point the polymer begins to dissolve. This type of polymer can be used to make reservoir or monolithic diffusion-controlled systems that degrade after their delivery role is complete. A final category of polymer has the active agent covalently attached by a labile bond to the backbone of a matrix polymer. When placed at the site of action, the labile bonds slowly degrade, releasing the active agent and forming a soluble polymer. The methods by which these concepts can be formulated into actual practical systems are illustrated in Fig. 6.

C. Osmotic Systems

Yet another class of delivery devices uses osmosis as the driving force. Osmotic effects are often a problem in diffusion-controlled systems because imbibition of water swells the device or dilutes the drug. However, several

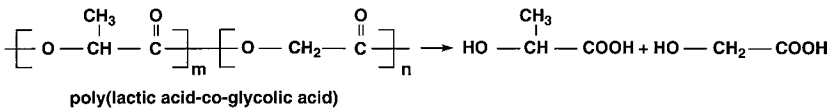


FIGURE 5 Biodegradation of poly(lactic acid-co-glycolic acid) to its simple acid precursors. The rate of biodegradation is a function of the copolymer composition. The pure homopolymers are both relatively crystalline and biodegrade slowly, but the more amorphous copolymers biodegrade more rapidly.

devices that actually use osmotic effects to control the release of drugs have been developed. These devices, called osmotic pumps, use the osmotic pressure developed by diffusion of water across a semipermeable membrane into a salt solution to push a solution of the active agent from the device. Osmotic pumps of various designs are widely applied in the pharmaceutical area, particularly in oral tablet formulations.

The forerunner of modern osmotic devices was the Rose–Nelson pump. Rose and Nelson were two Australian physiologists interested in the delivery of drugs to the gut of sheep and cattle. Their pump, illustrated in Fig. 7, consists of three chambers: a drug chamber, a salt chamber containing excess solid salt, and a water chamber.The drug and water chambers are separated by a rigid, semipermeable membrane. The difference in osmotic pressure across the membrane moves water from the water chamber into the salt chamber. The volume of the salt chamber increases because of this water flow, which distends the latex diaphragm separating the salt and drug chambers, thereby pumping drug out of the device.

The pumping rate of the Rose–Nelson pump is given by the equation

$$\frac{dM_t}{dt} = \frac{dV}{dt} \cdot c, \tag{10}$$

where dM_t/dt is the drug release rate, dV/dt is the volume flow of water into the salt chamber, and c is the concentration of drug in the drug chamber. The osmotic water flow across a membrane is given by the equation

$$\frac{dV}{dt} = \frac{A\theta\Delta\pi}{l}, \tag{11}$$

where dV/dt is a water flow across the membrane of area A, thickness l, and osmotic permeability θ ($cm^3 \cdot cm/cm^2 \cdot hr \cdot atm$), and $\Delta\pi$ is the osmotic pressure difference between the solutions on either side of the membrane. This equation is only strictly true for completely selective

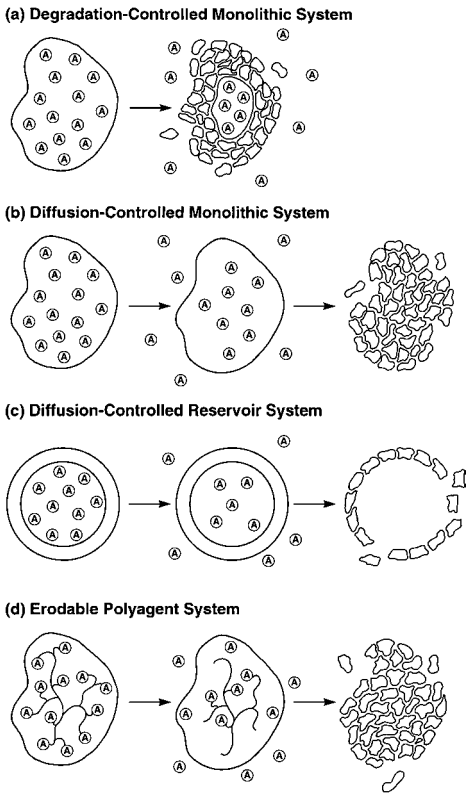


FIGURE 6 Various types of biodegradable drug delivery systems.

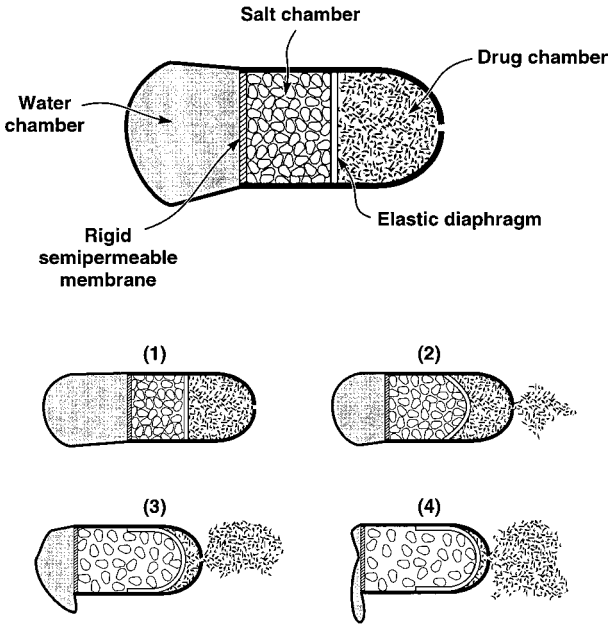


FIGURE 7 Mechanism of action of a Rose–Nelson osmotic pump, the precursor of today's osmotic controlled release delivery systems.

membranes, that is, membranes permeable to water but completely impermeable to the osmotic agent. However, this is a good approximation for most membranes. Substituting Eq. (11) for the flux across the membrane gives

$$\frac{dM_t}{dt} = \frac{A\theta\Delta\pi c}{l}. \qquad (12)$$

The osmotic pressure of the saturated salt solution is high, on the order of tens of atmospheres, and the small pressure required to pump the suspension of active agent is insignificant in comparison. Therefore, the rate of water permeation across the semipermeable membrane remains constant as long as sufficient solid salt is present in the salt chamber to maintain a saturated solution and hence a constant-osmotic-pressure driving force.

Variations of the Rose–Nelson pump have been developed as tools for drug delivery tests in animals. However, the development that made osmotic delivery a major method of achieving controlled drug release was the invention of the elementary osmotic pump by Theeuwes in 1974. The concept behind this invention is illustrated in Fig. 8. The device is a simplification of the Rose–Nelson pump, and eliminates the separate salt chamber by using the drug itself as the osmotic agent. The water required to power the device is supplied by the body, so a separate water chamber is also eliminated. The Theeuwes device is formed by compressing a drug having a suitable osmotic pressure into a tablet using a tableting machine. The tablet is then coated with a semipermeable membrane, usually cellulose acetate, and a small hole is drilled through the membrane coating. When the tablet is placed in an aqueous environment, the osmotic pressure of the soluble drug inside the tablet draws water through the semipermeable coating, forming a saturated aqueous solution inside the device. The membrane does not expand, so the increase in volume caused by the imbibition of water raises the hydrostatic pressure inside the tablet slightly. This pressure is relieved by a flow of saturated agent solution out of the device through the small orifice. Thus, the tablet acts as a small pump, in which water is drawn osmotically into the tablet through the membrane wall and then leaves as a saturated drug solution through the orifice. This process continues at a constant rate until all the solid drug inside the tablet has been dissolved and only a solution-filled shell remains. This residual dissolved drug continues to be delivered, but at a declining rate, until the osmotic pressures inside and outside the tablet are equal. The driving force that draws water into the device is the difference in osmotic pressure between the outside environment and a saturated drug solution. Therefore, the osmotic pressure of the dissolved drug solution has to be relatively high to overcome the osmotic pressure of the body. For drugs with solubilities greater than 5–10 wt%, this device functions very well. Later variations on the simple osmotic tablet design use water-soluble excipients to provide part of the osmotic pressure driving force; this overcomes the solubility limitation.

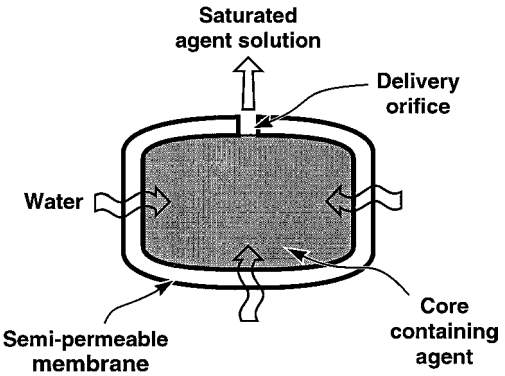


FIGURE 8 The Theeuwes elementary osmotic pump. This simple device consists of a core of water-soluble drug surrounded by a coating of a water-permeable polymer. A hole is drilled through the coating to allow the drug to escape.

III. IMPORTANT CONTROLLED RELEASE PRODUCTS

The controlled release sector is a rapidly expanding part of the pharmaceutical industry. Growth has occurred at the remarkable annual rate of 15% over the past decade, fueled by an explosion of new technologies. The value of the pharmaceuticals using controlled drug delivery reached $20 billion in 1999, and while only modest growth in the overall pharmaceutical market is projected for the next few years, the drug delivery share of the market is expected to continue to grow at least 15% per annum. As much as 20% of the U.S. pharmaceutical market is projected to be controlled release products by 2005.

The majority of drug delivery products reach the market as a result of a strategic alliance between a drug delivery company, which supplies the technology, and a pharmaceutical company, which supplies the drug and the resources needed for full development. A good example of such a collaboration is provided by protein and peptide drugs, an increasingly important area of pharmaceuticals driven by recent advances in biotechnology. Currently, a major factor limiting the use of these drugs is the need for their administration by repeated injections. This is because peptides and proteins are poorly absorbed in the GI tract, so cannot be delivered as oral tablets, and have very short biological half-lives, so cannot be delivered as a single, long-acting injection. Several specialized drug delivery companies are developing innovative techniques that will allow these drugs to be delivered orally, by nasal

inhalation, or in a long-acting injectable formulation. Alliances between these technology providers and the pharmaceutical companies producing the protein/peptide active agents increase the likelihood that these novel drug formulations will reach the marketplace.

Controlled drug delivery has a record of success in the pharmaceutical industry and can offer a high return on capital investment. Some relatively low-budget development programs, which have enabled existing, effective drugs to be administered to the patient by controlled release technology, have been very successful; examples include Lupron Depot (leuprolide, a hypothalamic releasing hormone used for the suppression of testosterone in the treatment of malignant neoplasms of the prostate), Procardia XL (nifedipine, a calcium channel blocker used for the treatment of hypertension and angina), and Cardizem CD (diltiazem, a calcium channel blocker with properties similar to nifedipine). These are all billion-dollar products that employ controlled release or delivery techniques. To develop a new chemical entity through to regulatory approval in the mid-1990s took, on average, 10–12 years and cost about $300–600 million. In contrast, an existing drug can be reformulated into an innovative drug delivery system in 5–7 years at a cost of about $20–$100 million. Interestingly, there are also recent examples in which controlled release is no longer a simple reformulation of an old drug. For example, new drugs are being developed and marketed for the first time as controlled release products. Some of these drugs might not have reached the market except for controlled release technology—felodipine and omeprazole are examples.

A. Oral Formulations

The oral route is by far the most common and convenient method of delivering drugs to the body. Unfortunately, the method has a number of problems that interfere with effective drug delivery. First, a drug taken by mouth is immediately exposed to low-pH stomach acids containing high concentrations of digestive enzymes. Many drugs are chemically degraded or enzymatically metabolized in the stomach before they are absorbed. Drugs that are absorbed then enter the portal circulation and may be destroyed by the first-pass metabolism in the liver described earlier. Controlled release is a method of avoiding these problems.

The typical transit time of material through the GI tract is 12–18 hr, so most controlled release oral formulations are designed to deliver their loading of drug over a 6- to 15-hr period. In this way the action of short-half-life drugs or rapidly absorbed drugs can be spread over a prolonged period. Drugs that might require dosing two or three times a day to achieve relatively uniform and nontoxic blood levels can then be dispensed as a single once-a-day tablet. This simple change, by improving patient compliance and producing a more controlled constant blood level, has produced measurable improvements in efficacy and reduced toxicity for many drugs. Many oral controlled release formulations are designed to produce a relatively low drug delivery rate for the first 1–3 hr while the formulation is in the stomach, followed by prolonged controlled release of the drug once the formulation has reached the GI tract. This avoids chemical degradation of the drug in the aggressive environment of the stomach. This type of delivery is, for example, particularly important for polypeptide drugs which are rapidly and completely destroyed if delivered to the stomach. Delivery to the GI tract is also done to achieve local delivery of the drug, such as the anti-inflammatory Mesalazine for irritable bowel disease and ulcerative colitis.

The precursors of today's controlled release oral formulations were enteric tablets based on various wax matrices designed to circumvent degradation in the stomach. Enteric formulations were later improved by using new, more reliable polymers. By the mid-1970s, the first oral controlled drug delivery systems began to appear. Two important delivery technologies developed at that time were Alza's Oros osmotic controlled release system and Elan's Sodas multiparticulate system. Elan's Sodas system consisted of large numbers of micropellets, each designed to release a microdose of drug by diffusion from a matrix at a predetermined rate. By blending pellets with different release profiles, the overall target rate was achieved. Since then, a wide variety of other oral formulations using osmosis and diffusion have been produced, as well as slow-release bioerodible tablets, ion exchange beads, multiple-layer tablets, and others.

If the drug is relatively water-soluble, osmotic or simple table formulations are often used to achieve controlled delivery. However, with more-insoluble drugs, release of the complete dosage from a single tablet in an 8- to 12-hr period may be difficult. For such drugs, a microencapsulated or granulated form of the drug is enclosed in a gelatin capsule. Microencapsulation exposes a much greater surface area of the device to interact with the body, so drugs that dissolve and diffuse slowly can still be completely released in an 8- to 12-hr period. Drugs can be microencapsulated by physical and chemical methods. Physical methods include encapsulation by pan coating, gravity flow, centrifugation, and fluid bed coating. Chemical microencapsulation normally involves a two-step process called coacervation. Drug particles or droplets of drug solution are first suspended in a polymer solution. Precipitation of the polymer from solution is then caused by, for example, changing the temperature or adding a nonsolvent. The polymer then coats the drug particles to form the microcapsule.

The leading developers of oral drug delivery formulations are Alza and Elan. Other important producers are Skyepharma, which has developed a technology called Geomatrix (a multilayer tablet with each layer releasing the drug at a different rate), R. P. Scherer, which has several different technologies including Zydis (a lyophilized tablet), and Eurand, with several technologies for controlled release, taste masking, and improved bioavailability.

B. Transdermal Systems

Scopolamine, for control of motion sickness, was the first drug to be marketed in the form of a transdermal patch system. Since then the market for transdermal patches has grown steadily. However, the number of drugs that can be delivered through the skin is more limited than was anticipated when the first patches were introduced in the 1980s. The main problem limiting widespread use is the low permeability of most drugs through the skin. Depending on the drug, skin permeabilities are in the range 0.01–10 μg of drug/cm$^2 \cdot$ hr. Because the maximum acceptable size of a transdermal patch is limited to about 50 cm^2, drugs delivered through the skin must be effective at doses of 0.01 mg/day for a poorly skin-permeable drug and 10 mg/day for a highly skin-permeable drug. Very active skin-permeable drugs are required to make transdermal drug delivery possible. Despite the enormous interest in transdermal delivery in academia and industry, the number of drugs delivered transdermally is currently limited to nitroglycerin, nicotine, estradiol, clonidine, fentanyl, testosterone, and isorbide nitrate. Nitroglycerin and nicotine are currently the most important drugs, but the area of hormone replacement therapy is growing as a result of improved estradiol and testosterone delivery patches.

The three main types of diffusion-controlled transdermal devices are shown schematically in Fig. 9. The simple adhesive device (Fig. 9, top) has a two-layer "Band-aid" configuration comprising the backing layer coated with adhesive. The drug is mixed in the adhesive layer used to fix the bandage to the skin. These medicated bandages bring a known quantity of drug to a known area of skin for a known period of time, but have no mechanism for controlling the rate at which the drug is delivered to the patient.

The second type of device is a monolithic system (Fig. 9, middle) incorporating a backing layer, a matrix layer, and an adhesive layer. The matrix layer consists of a polymer material in which the solid drug is dispersed; the rate at which the drug is released from the device is controlled by this polymer matrix. With this type of system, the drug release rate falls off with time as the drug in the skin-contacting side of the matrix is depleted.

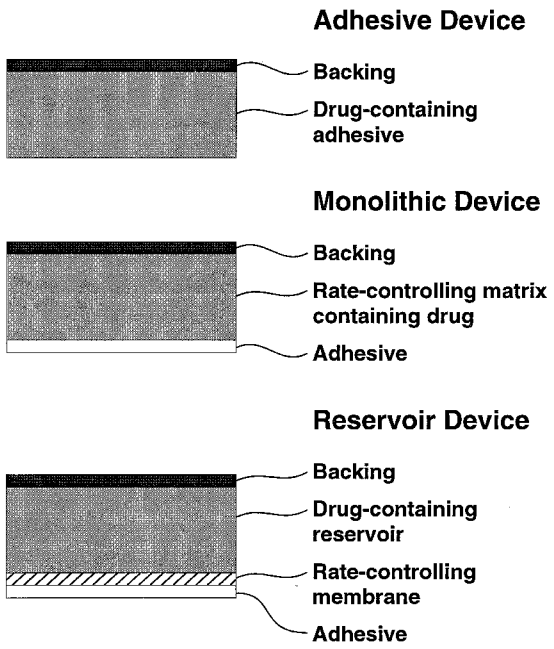


FIGURE 9 Transdermal patch designs.

The third type of device is the reservoir system (Fig. 9, bottom). In this case, the drug—usually in liquid or gel form—is contained in a reservoir separated from the skin by an inert membrane that controls the rate at which drug is delivered to the skin. These devices offer an important advantage over the monolithic geometry: as long as the drug solution in the reservoir remains saturated, the drug release rate through the membrane is constant.

The pattern of drug release from the device is important. If drug is delivered to the skin at less than the maximum rate at which it can be absorbed, the device is the primary dosage-controlling mechanism. When the drug is delivered faster than the skin can absorb it, the skin surface is then saturated with drug at all times, and the limiting factor for systematic dosage is the rate of absorption through the skin. Thus, at least in principle, devices for which the dosage-controlling mechanism is either the skin or the device can be designed.

To reach the systemic blood circulatory system, drug from a transdermal patch must cross several layers of skin, as shown in Fig. 10. The top surface layer of skin, called the *stratum corneum*, represents the main barrier to drug permeation. The stratum corneum is only 10–15 μm thick, but it consists of layers of flattened, cornified cells that are quite impermeable. The interspace between these cells is filled with lipids, giving the structure a "bricks-and-mortar" form, with the cells being the bricks and the lipids being the mortar. The most important pathway for drug absorption is through the lipid (mortar), which dictates

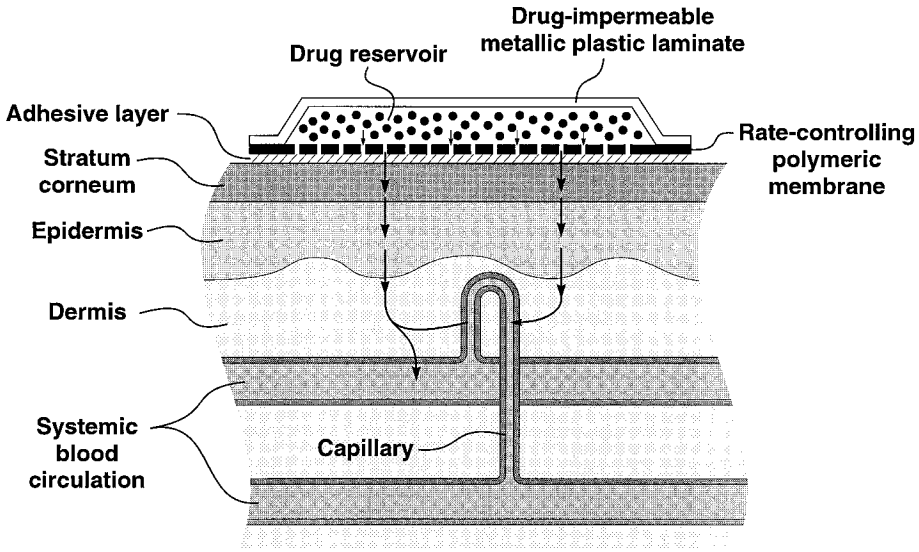


FIGURE 10 Membrane-moderated transdermal drug delivery system (not to scale). Drug permeates from the patch reservoir through the protective outer layers of the skin and is absorbed into the underlying capillaries of the general systemic blood circulation.

the characteristics of usefully permeable drugs. These are drugs with a molecular weight of less than 1000 dalton, a low melting point, an octanol/water partition coefficient between 0 and 2, and few polar centers.

Many drugs do not have the chemistry required to achieve high skin permeabilities, so various methods of enhancing drug permeability have been developed. One method is to soften and swell the stratum corneum by dissolving or dispersing the drug in a simple solvent. For example, the first estradiol-delivery transdermal patch used ethanol to enhance the skin's permeability to estradiol. In the absence of ethanol, estradiol flux through the skin is very low, on the order of 0.01 $\mu g/cm^2 \cdot hr$. However, if a suspension of estradiol in ethanol is applied to the skin, the estradiol flux increases 10- to 20-fold. A second way to enhance skin permeability is to use a compound such as a long-chain fatty acid, ester, or alcohol. These compounds penetrate the stratum corneum more slowly than small molecules such as ethanol but have a more prolonged plasticizing effect. A third method combines elements of the first two. An enhancer formulation containing both a simple solvent and a fatty component is used to combine the rapid onset of solvent effect with the prolonged action of the fatty component.

Another method of enhancing drug permeation is to increase the driving force for drug permeation by using a small electric current. This last approach, called iontophoresis, has been widely studied. The principle of iontophoresis is illustrated in Fig. 11. In this method, a battery is connected to two electrodes on the skin. An ionized drug placed in contact with one electrode will migrate under the influence of the voltage gradient through the skin and enter the system circulation; very large enhancements can be obtained. The earliest devices having the essential features of iontophoresis date back to the 1890s, although apparently their objective was to shock their subjects rather than to administer drugs to them. The first modern device appeared in 1972, and advances in electronics have since allowed smaller and smaller devices to be built. The newest devices have a built-in battery layer and are comparable in size to a normal transdermal patch. Iontophoresis appears to be a particularly promising tool for the delivery of very active peptides, small proteins, or oligonucleotides, which are otherwise almost completely

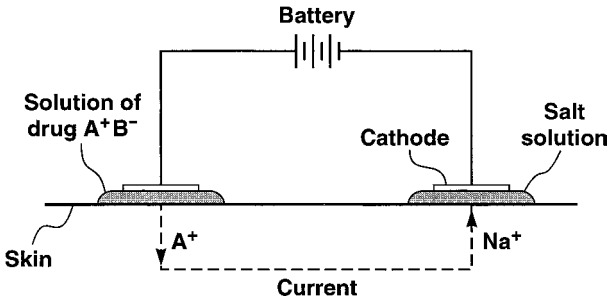


FIGURE 11 Mechanism of an iontophoresis patch. The poles of a small battery are connected to the skin. A solution of an ionized drug placed at one electrode migrates through the skin into the systemic circulation under the action of the voltage driving force of the battery.

skin-impermeable. Under the action of a small voltage gradient, the skin permeation rate of the drug increases 10- to 100-fold. Another advantage of iontophoresis is that the voltage driving force can be easily regulated; this allows pulsatile drug delivery or variable delivery according to the patient's needs. This control is useful in the delivery of analgesics such as fentanyl for the control of pain. Despite the many patents in this field, no commercial products have yet reached the market.

The main advantages of transdermal over oral drug delivery are the ability to maintain constant plasma levels with short-half-life drugs and to avoid the hostile conditions of the gastrointestinal tract and consequent drug deactivation because of the hepatic first-pass effect. Moreover, patient compliance tends to increase, reducing the number of administrations required. Patches are now available for once-a-day, twice-a-week, and once-a-week treatment.

C. Nasal Spray/Inhalers

Drug delivery by inhalation has a long history and is an obvious method of administering agents that act on the respiratory system. However, it is now being used increasingly to deliver systemically active drugs. The first hand-held, pressurized, metered inhaler was launched in 1960; several other products have been introduced since. Intranasal delivery is currently employed in treatments for migraine, smoking cessation, acute pain relief, nocturnal enuresis, osteoporosis, and vitamin B_{12} deficiency. In 1999, Aviron's intranasal influenza vaccine, FluMist, was filed with the FDA, and several other vaccines are being considered for nasal administration. Other applications for nasal delivery under development include cancer therapy, epilepsy control, antiemetics, and treatment of insulin-dependent diabetes.

The advantages of the nasal cavity over other drug administration routes are rapid, direct access to the systemic circulation and complete avoidance of first-pass metabolism. The nasal cavity is also a far less aggressive environment than the gastrointestinal tract, and so is particularly useful for the delivery of peptides and proteins, which are easily degraded in the stomach. Patient compliance is also improved, particularly if the alternative treatment is intravenous injection. However, the nasal route is not without its problems. Issues that need to be considered are the rate of drug absorption through the nasal mucosa, the residence time of the formulation at the site of delivery, local toxicity and tolerability, and degradation of the drug in the nasal cavity. The rate of absorption of the drug can be controlled by including carriers and enhancers to modify permeation, or by chemical modification of the drug into a more readily absorbed form. Most nasal devices deliver the drug as a fine liquid or solid spray; the average particle size significantly affects the rate of drug absorption. Intranasal delivery of systemic drugs requires sophisticated delivery devices to ensure accurate, repeatable dosing and minimal irritation. Dosing accuracy is particularly important for delivery of potent and expensive drugs such as the new generation of peptide and protein products. Long-term nasal delivery may also require dose-counters and lock-out systems to prevent overdosing as, for example, in the pain relief area.

Prior to 1985, chlorofluorocarbons were widely used as inert nontoxic inhaler propellants; these compounds are now all but banned. Consequently, much recent research has centered on the development of dry powder inhalers, which deliver the active ingredient as ultrafine particles directly to the lungs with minimal deposition in the mouth and trachea. Some of these devices are activated directly by the inspiration of the patient without the need to coordinate activation and inspiration.

Although dry powder inhalers have a number of therapeutic benefits, they also have problems. For example, contact of the powder formulation with moisture can lead to agglomeration and inaccurate dosing, and dose uniformity is hard to achieve. The potential for nasal delivery is greatest in two areas: local delivery to the lung for respiratory treatment diseases and systemic delivery of a broad variety of drugs via the alveoli, which have highly absorptive properties. Its greater long-term potential is in the delivery of macromolecules, but further research is needed to determine the long-term immunogenicity, reproducibility, and stability of the delivery systems.

D. Targeted Drug Delivery

Targeted or directed drug delivery is a relatively new method of delivering drugs. The objective is to alter the effectiveness of drugs by targeting their delivery directly to the site needing drug action. Promising techniques include the use of liposomes, polyethylene glycol (PEG)-coated molecules, blood–brain barrier transfer agents, and several antibody conjugate approaches.

The most advanced targeted delivery technology uses liposomes, which are ultrafine water/oil/water emulsions in which the emulsion droplets consist of lipid vesicles containing an aqueous drug solution. The surface of the vesicles is sometimes modified to promote targeting of the lipid-drug-containing vesicle to selected tissues. The first liposomes were developed in 1965 as a model of biological membranes. Their potential as a drug delivery system was recognized later; now, after many years of gestation, liposomes are finally being introduced in commercial

products. The mean diameter of liposomes is less than 0.1 μm. This allows them selectively to extravasate into tissues characterized by leaky vasculature, such as solid tumors, achieving targeted delivery to the diseased organ with low adverse effects on normal tissues.

The first liposome product on the market was Ambisome, containing amphotericin B, an antifungal encapsulated in liposomes to reduce its toxicity. Since then, liposomal preparations of other drugs have been developed, most importantly Daunoxome, an anticancer product containing the antitumoral drug daunorubicin for the treatment of Kaposi's sarcoma. Liposomes are also being investigated as vehicles for gene therapy. Their main advantages over viral carriers is a higher loading capacity and a lower risk of evoking an immune response. Surface-coating liposomes with antibodies to allow active targeting to a particular disease site is also being studied in clinical trials. If these products are successful, targeted drug delivery will be a breakthrough in the treatment of a number of major diseases.

A second approach to targeted drug delivery uses polyethylene glycol (PEG), a water-soluble polymer covalently linked to proteins, which alters their properties and extends their potential use. The main product using this approach is for α-interferon for the treatment of hepatitis C. This technique has been applied to several proteins including adenosine deaminase, cytokines, and granulocyte-macrophage colony-stimulating factor. Generally the results obtained by the modification with PEG are increased circulating life, reduced immunogenicity and antigenicity, and increased stability and solubility with a minimal loss of biological activity.

E. Implants

Three types of implantable drug delivery devices are currently in use or being developed. The first type is an implantable polymeric capsule most commonly made from silicone rubber, which when placed under the skin delivers the drug load at a constant rate for as long as 2–5 years. Upjohn's Depo-Provera and American Homes' Norplant, used to deliver contraceptive steroids for long-term contraception, are examples of this approach. Implantable devices achieve long-term, controlled delivery of the drug and patient compliance is no longer an issue, both significant advantages. The key disadvantage is the size of the device, which means minor surgery is required to insert the implant and later to remove the device once its drug delivery role has been completed. There is also a potential to release all of the drug in a few days if the capsule begins to leak. For these reasons this type of nondegradable device has not become a major product.

A second type of device uses implants made from biodegradable polymers. Because the device is biodegradable, device retrieval once drug delivery is complete is no longer necessary. This makes the device much more acceptable to patients. Unfortunately, it is technically difficult to create implantable, biodegradable devices that deliver drug reliably for more than 1 or 2 months. Therefore, these devices are generally restricted to delivering drugs for 4–6 weeks. Examples include Abbot's Lupron Depot formulation, a single monthly injection of leuprolide for endometriosis, and Zoladex, Zenaca's 4-week implantable formulations of goserelin for endometriosis in women and prostrate cancer in men.

A final category of implantable devices, in late-stage development, is miniature pumps driven by osmosis or fluorocarbon propellants. These pumps are designed to deliver essentially any drug at a predetermined rate for weeks or even months. The problem to be solved in this case is to make the pump small enough and reliable enough that the trauma of device removal is outweighted by its drug delivery benefits.

IV. FUTURE DIRECTIONS

The era of modern controlled drug delivery started with the launching of the first product registered at the U.S. Food and Drug Agency in terms of both the total amount of drug delivered and the rate of drug delivery. This first product, Alza's Ocusert, was launched in 1974. By 1990 the total controlled release market was approximately $1 billion. Since then the market has grown 20-fold, and the number of such products is expected to continue to grow rapidly in the next few years.

The largest growth area will be the extension of already developed controlled release technologies to a wider number of drugs. A long list of oral and some transdermal controlled release products are in late-stage clinical trials and should appear on the market in the next few years.

A second area of significant future growth will be the use of controlled release systems to deliver the new generation of peptide and protein drugs. As a consequence of the sequencing of the human genome, the information required to design protein and peptide drugs to treat important diseases is at hand. However, many of these new drugs are essentially ineffective if administered as conventional pharmaceutical formulations. New, well-designed controlled release systems that can deliver the drugs intact at a prolonged steady rate close to the site of action are required; such systems are being actively developed.

A third potential growth area for controlled release technology is the development of systems for targeted drug delivery. Drug delivery to the ultimate site of action is a multistep process. Conventional controlled drug delivery systems generally only address the first step, the rate of delivery of drug to the body. The path of the drug from the dosage site to the diseased cells is largely uncontrolled. Targeted drug delivery systems attempt to improve this step in the delivery process, in effect, to improve the aim of Ehrlich's magic bullet. Antibody-coated liposomes and the PEG-peptide products described above are the first simple prototype product designs to tackle this problem. Over time, more sophisticated and more effective techniques will be developed.

Finally, all the controlled release products described in this article are preprogrammed, with the rate of delivery being established by the designer of the device. No subsequent adjustment of the drug delivery rate in response to patient need is possible. However, many disease conditions are episodic, for example, diabetes, stroke, migraines, heart attacks, and epilepsy. Controlled drug delivery systems that could sense the body's need and automatically deliver the drug at the rate required would be a huge benefit. In the intensive care facilities of modern hospitals, this type of control is achieved through continuous electronic sensors and monitoring by the attending nurses and physicians. In the future, development of microelectronic/micromechanical patient-portable machines to produce the same level of control on ambulatory, at-risk patients can be imagined.

SEE ALSO THE FOLLOWING ARTICLES

AEROSOLS • BIOPOLYMERS • ELECTROPHORESIS • ION TRANSPORT ACROSS BIOLOGICAL MEMBRANES • MEMBRANES, SYNTHETIC, APPLICATIONS • PHARMACEUTICALS • PHARMACOKINETICS

BIBLIOGRAPHY

Baker, R. (1987). "Controlled Release of Biologically Active Agents," Wiley, New York.

Benita, S. (1996). "Microencapsulation: Methods and Industrial Applications," Marcel Dekker, New York.

Chasin, M., and Langer, R. (1990). "Biodegradable Polymers as Drug Delivery Systems," Marcel Dekker, New York.

Chien, Y., Su, K., and Chang, S. (1989). "Nasal Systemic Drug Delivery," Marcel Dekker, New York.

Clark, A. (1995). "Medical aerosol inhalers: Past, present and future," *Aerosol Sci. Technol.* **22,** 374–391.

Friend, D. (1992). "Oral Colon-Specific Drug Delivery," CRC Press, Boca Raton, Fl.

Katre, N. (1993). "The conjugation of proteins with polyethylene glycol and other polymers," *Adv. Drug Deliv. Rev.* **10,** 91–114.

Lasic, D., and Papahadjopoulos, D. (1998). "Medical Applications of Liposomes," Elsevier Science, New York.

Potts, R., and Guy, R. (1997). "Mechanisms of Transdermal Drug Delivery," Marcel Dekker, New York.

Santus, G., and Baker, R. (1995). "Osmotic drug delivery: A review of the patent literature," *J. Controlled Release* **35,** 1–21.

Smith, E., and Maibach, H. (1995). "Percutaneous Penetration Enhancers," CRC Press, Boca Raton, FL.

Wise, D. (2000). "Handbook of Pharmaceutical Controlled Release Technology," Marcel Dekker, New York.

Physical Organic Chemistry

Charles L. Perrin
University of California, San Diego

GLOSSARY

Aromaticity Unusual stability of cyclic conjugated systems containing $4n + 2$ pi electrons.
Brønsted relationship Linearity between logarithms of the catalytic constants for acid or base and pK.
Chiral Not superimposable upon its mirror image.
***Cis/trans* isomerism** Isomerism arising from sidedness on a double bond or ring.
Conformations Structures interconvertible by rotation about single bonds.
Diastereomers Stereoisomers that are not mirror images of each other.
Enantiomers Stereoisomers that are (nonsuperimposable) mirror images of each other.
General acid catalysis Catalysis where each acid present functions as a catalyst, rather than H_3O^+ alone.
Hammett equation Quantitative comparison with substituent effects in benzoic acid dissociation.
Intermediates Energy minina between reactant and product on a reaction coordinate.
Linear free energy relationship Logarithmic linearity between rate and equilibrium constants.
Nucleophile Atom or reactant that uses an electron pair to form a bond to an electrophile.
Orbital One-electron wave function.
Order (kinetic order) Power of the concentration of a chemical species entering into a rate equation.
Pericyclic reaction Reorganization of electrons around a cycle of nuclei, with simultaneous bond breaking and bond making.
Product-development control Predominant formation of the more stable product because its transition state partakes of that stability.
Racemic mixture A 50:50 mixture of two enantiomers.
Reaction coordinate Geometric parameter that measures progress from reactants to products.
Stereoisomers Different molecules with the same connectivity (bonding relationships between atom pairs).
Torsional strain Destabilization of eclipsed conformations.
Transition state Position of maximum energy along reaction coordinate.

PHYSICAL ORGANIC CHEMISTRY undertakes the investigation of the phenomena of organic chemistry by quantitative and mathematical methods. This was the original approach, as presented in L. P. Hammett's influential book, "Physical Organic Chemistry: Reaction Rates, Equilibria, and Mechanisms," published in 1940. Prior to Hammett's studies, organic chemistry had largely been viewed as a collection of empirical observations, without any underlying sense. Hammett's contribution was to take the methodology of physical chemistry, apply it to organic chemistry, and derive regularities that placed organic reactivity on a quantitative basis.

Since then, physical organic chemistry has broadened its focus to the structure, properties, and reactions of organic molecules, and especially to the relationship between structure and reactivity. Physical organic chemistry asks how chemical reactivity depends on molecular structure. Part of the answer comes from detailed and quantitative studies of some reactions. Part comes from recognizing analogies among classes of reactions. This latter is a particularly powerful method since it permits the extension of understanding from one well studied class of reactions to another, less well understood one. The underlying principle is that small perturbations of molecular structure are unlikely to lead to major changes in reactivity. This principle does not always hold, but it holds often enough to provide a broad general theory of chemical reactivity. In recent years physical organic chemists have been successful in codifying the underlying principles of chemical structure and reactivity. The results have provided not only a deep understanding of organic reactions, but also applications to organic synthesis, biochemical processes, and materials science.

I. MOLECULAR STRUCTURE: BONDING

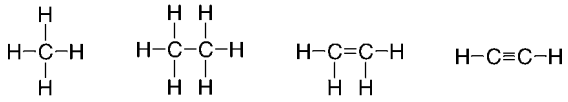
A. Covalent Bonding

Since structure determines chemical reactivity, it is necessary to have a thorough understanding of molecular structure. Two atoms are held together by the sharing of valence electrons between them. The interaction that arises from sharing a pair of electrons is called a covalent bond. In early illustrations the two electrons were shown as dots. Now the covalent bond is symbolized by a line connecting the two atoms. The simplest case is the hydrogen molecule (**1**), formed from two hydrogen atoms, each with one valence electron. The line symbolizes the two electrons that are shared.

H–H

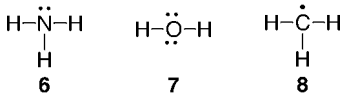
1

A covalent bond represents a stabilization of the molecule relative to the separated atoms. Through sharing of electrons, each atom achieves an "octet configuration" analogous to that of the noble gases. For example, each of the hydrogen atoms in **1** is associated with two electrons, just as in a helium atom. Thus the hydrogens achieve the inertness or unreactivity of a noble gas. Hydrogen and helium are unique in that only two electrons are necessary to complete a filled shell. For the next rows of the periodic table, eight valence electrons are necessary, hence the designation "octet." For example, carbon in methane (**2**) achieves its octet by sharing its four valence electrons with four hydrogens, each contributing another electron, and each of the carbons in ethane (**3**) shares a pair of electrons with each of three hydrogens and also with each other. In ethene (**4**) each of two carbons achieves its octet by sharing two pairs of electrons with each other and also two additional electron pairs with two hydrogens. The four electrons that are shared between the two carbons form a double bond, symbolized by two connecting lines. Similarly, ethyne (**5**) has a triple bond, with six electrons shared between the two carbons, each of which achieves an octet.



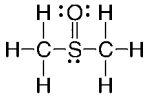
2 **3** **4** **5**

A system of graphic symbols, called Lewis structures, permits the illustration of molecular structures. According to this system, every valence electron must be shown. Some valence electrons are in covalent bonds, whether single, double, or triple. Some valence electrons are unshared and are symbolized as dots on the atoms to which they belong. Sometimes these are paired, as in ammonia (**6**) or water (**7**), where the unshared electrons are called lone pairs. Sometimes an unshared electron is unpaired, as in methyl radical (**8**), where it is called an odd electron. This last is unusual in that the carbon lacks an octet. In nearly all stable molecules every atom has its octet, and nonoctet atoms are usually quite reactive (NO and NO_2 are exceptions).

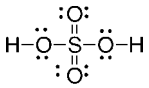


6 **7** **8**

Exceptions to the octet rule are possible. Metallic elements, at the left of the periodic table, often lack an octet. For example, the boron in BF_3 has only six electrons. More common exceptions are with atoms in the second full row of the periodic table, such as the phosphorus in PCl_5 or the sulfurs in dimethyl sulfoxide (**9**) and sulfuric acid (**10**), which have 10, 10, and 12 electrons, respectively.



9

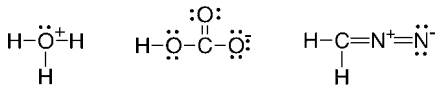


10

A further feature of Lewis structures is the designation of formal charge at each atom. The formal charge Z_f is defined as follows, where N_{val} is the number of valence electrons, N_{bond} is the total number of bonds in which the atom is engaged, N_{pair} is the number of lone pairs on the atom, and N_{odd} is the number of its odd electrons (almost always zero):

$$Z_f = N_{val} - N_{bond} - 2N_{pair} - N_{odd}. \qquad (1)$$

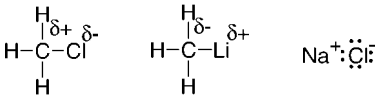
This equation reflects the loss or gain of electrons on forming covalent bonds, since N_{val} is the number of electrons that the atom originally had, $2N_{pair} + N_{odd}$ electrons remain on the atom, and the two electrons in any bond are considered to be shared equally, so that only one of them "belongs" to the atom. For example, the formal charges in hydronium ion (**11**), bicarbonate ion (**12**), and diazomethane (**13**) are as shown (omitted for atoms with $Z_f = 0$), since $Z_f(O_{\mathbf{11}}) = 6 - 3 - 2 \times 1 = +1$, $Z_f(O_{right}) = 6 - 1 - 2 \times 3 = -1$, $Z_f(O_{other}) = 6 - 2 - 2 \times 2 = 0$, $Z_f(N_{middle}) = 5 - 4 = +1$, and $Z_f(N_{right}) = 5 - 2 - 2 \times 2 = -1$.



11 **12** **13**

The assumption of equal sharing of electrons in a covalent bond is an oversimplification for the purpose of assigning formal charge. In fact, the electrons are held more tightly by whichever atom is more electronegative. Electronegativity is an elusive concept that is best treated qualitatively. Electronegativity increases toward the right of the periodic table (C < N < O < F, Si < P < S < Cl) and toward the top (I < Br < Cl < F, Se < S < O). Thus fluorine is the most electronegative and the other halogens are also electronegative since they need only one additional electron to complete their octet. Metals are not electronegative, and both carbon and hydrogen are intermediate and of nearly equal electronegativity. In cases of unequal sharing, covalent bonds have a polar character, with partial positive and partial negative charges symbolized by $\delta+$ and $\delta-$, as in methyl chloride (**14**) and methyl lithium (**15**). In the extreme case of a large difference in electronegativities, there may be no sharing. Instead the electron is fully transferred and the bond is said to be ionic, as in sodium chloride (**16**). The imbalance of charge leads to a local dipole moment μ given by the following equation, where q is the partial (or full) charge and d is the distance between the positive and negative charges:

$$\mu = qd. \qquad (2)$$

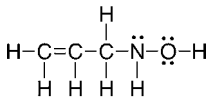


14 **15** **16**

It is unwieldy to display each hydrogen in a Lewis structure. Often, condensed structures are used, with the n hydrogens on an atom written as H_n following (or sometimes preceding) the atom, and with single bonds omitted and instead merely implied by juxtaposing the bonded atoms. Thus methyl chloride (**14**) and methyl lithium (**15**) become CH_3Cl and CH_3Li, respectively. Also, parentheses are used to indicate branching, as in $CH_3C(=O)CH_3$ or $CH_3CH(OH)CH_3$, but they may be omitted, as in CH_3COOH, which can be misleading.

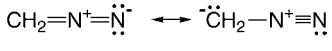
B. Resonance

For many molecules the Lewis symbolism provides an adequate representation of the molecular structure. The general requirement for such molecules is that they must have only single bonds and lone pairs, and they also must have neither nonoctet atoms nor conjugated multiple (double or triple) bonds. A conjugated multiple bond is one that is adjacent to a lone pair or multiple bond, meaning that it is separated by only one bond. For example, **17** satisfies the requirement, but **12** and **13** do not, since they have a lone pair (on oxygen or nitrogen) adjacent to a C═O (vertical) or C═N (separated by C─O or N═N).



17

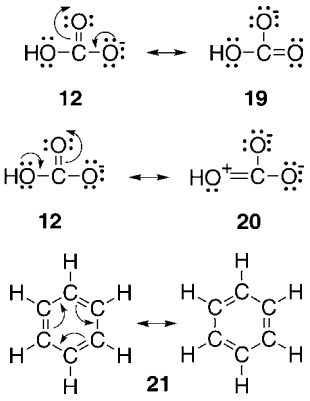
Molecules for which this requirement is not met have delocalized electrons, electrons that cannot be assigned to a covalent bond between two atoms or to a lone pair or odd electron on a single atom. One way to describe molecules with delocalized electrons is through the symbolism of resonance. The phenomenon is quantum mechanical and not readily understood in classical terms. A resonance hybrid is composed of several resonance forms each of which is a single Lewis structure with localized electrons. The individual resonance forms have no independent existence, except that they together contribute to the hybrid. The hybrid is symbolized with a double-headed arrow associating the individual structures with each other. For example, diazomethane is a hybrid of the previous resonance form **13** and another one, **18**.



13 **18**

There are three rules to be obeyed in dealing with resonance: (1) In each contributing resonance form every atom must satisfy the octet rule (as modified by the exceptions above) and resonance forms with a nonoctet atom can usually be ignored; (2) resonance expresses the delocalization of electrons, but the resonance hybrid has a single geometry, with fixed nuclei; and (3) resonance provides stabilization, such that the resonance hybrid is of lower energy than would be expected for any of its contributing resonance forms. This stabilization is called resonance energy.

Electron pushing, symbolized with curved arrows, provides a convenient method for generating additional resonance forms. The rule is to delocalize a lone pair (or a pair of electrons in a multiple bond, or, more rarely, in a single bond) toward an adjacent atom so as to form a multiple bond, while also removing an electron pair from that adjacent atom to avoid a violation of the octet rule. For example, delocalizing an electron pair from O_{right} of bicarbonate ion (**12**) generates another resonance form (**19**), and delocalizing an electron pair from O_{left} generates yet another (**20**). Benzene (**21**) is an example of delocalizing electrons in multiple bonds. The second resonance form is not just the first one rotated by 60°; the nuclei remain fixed, and the electrons of the double bonds are delocalized.

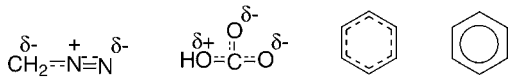


12 **19**

12 **20**

21

It is tedious to draw all resonance forms, and a shorthand notation is often used. Bonds that are present in some resonance forms but not in all are shown as dotted, formal charges that are present in some resonance forms but not all are shown as $\delta+$ or $\delta-$ (same symbolism as before, different context), and lone pairs are not shown. Thus diazomethane (**13**, **18**), bicarbonate ion (**12**, **19**, **20**), and benzene (**21**) become **22**, **23**, and **24**, respectively. For this last a further abbreviation, often applied to cyclic compounds, is to omit hydrogens attached to carbons and to symbolize each carbon as a vertex. Moreover, the six dotted lines are often simplified to a circle, as in **24′**.



22 **23** **24** **24'**

C. Atomic Orbitals

Lewis structures do not describe the geometry of molecules. The set of bonds in a molecule describes only the connectivity—which atom is bonded to which other atom or atoms. Indeed, most of the structures **2**–**20** are not drawn with any attempt at correct geometry. To describe the geometry, another approach is necessary. Orbital theory provides a complementary approach to molecular structure.

Orbital theory takes account of the wave nature of electrons. The uncertainty principle of quantum mechanics prohibits locating an electron exactly. Instead only probabilities can be specified, as described by a wave function ψ. For a single electron that wave function is called an orbital, and its square is the probability of finding the electron "at" the point x, y, z in three-dimensional space:

$$P(x, y, z) = [\psi(x, y, z)]^2. \tag{3}$$

It is possible to express ψ as a mathematical formula with a value at every point in space but it is often more convenient just to sketch surfaces of constant ψ. For some orbitals, called s orbitals, the surface is a sphere. For some other orbitals, called p orbitals, the surface resembles a pair of flattened spheres separated by a plane where $\psi(x, y, z) = 0$. Such a plane is called a nodal plane and it divides the region of space where $\psi(x, y, z) > 0$ from the region where $\psi(x, y, z) < 0$. There is no further significance to positive or negative $\psi(x, y, z)$ since only the probability in Eq. (3) is measurable. There are also d orbitals, with two nodal planes.

It takes a little imagination to reconstruct a three-dimensional surface of constant ψ from its illustration on a two-dimensional page. The sphere of an s orbital can be illustrated as a circle, as shown in Fig. 1. The near-spheres of a p orbital are rarely illustrated accurately. Instead stylized (*very* approximate) versions are often used, as also shown in Fig. 1.

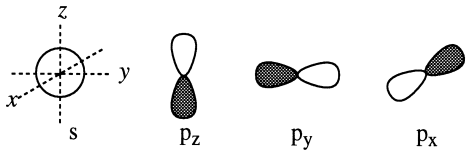


FIGURE 1 Atomic orbitals.

D. Molecular Orbitals

For molecules the orbitals become even more complicated. In general, they are not simply s or p orbitals on a single atom, but orbitals that spread over the entire molecule. In some cases they can be approximated in terms of two atomic orbitals on two adjacent atoms. For example, in the simplest case of the hydrogen molecule, the molecular orbital ψ_{MO} can be expressed as a linear combination ("sum") of the s orbitals on each of the two atoms, labeled A and B, as follows (where c_A and c_B are numerical coefficients that can be evaluated by quantum mechanical calculation):

$$\psi_{MO} = c_A\psi_A + c_B\psi_B. \quad (4)$$

According to the Pauli exclusion principle, only two electrons are permitted in this or any molecular orbital.

This molecular orbital could be illustrated by drawing its surface of constant ψ. It is more common just to show the stylized surfaces of the two atomic orbitals and imagine that the surface of the molecular orbital is the boundary envelope of those atomic orbitals, as suggested in Fig. 2. If the probability, Eq. (3), associated with ψ_{MO} is evaluated, it is found that the two electrons in this molecular orbital are more likely to be found in the overlap region, the region of space between the two nuclei where the atomic orbitals overlap. Since this is a region where the electrons are attracted to both positively charged nuclei, the probability increase represents a stabilization. This is the molecular orbital counterpart of the stabilization associated with each hydrogen's achievement of its "octet."

Molecular orbitals [Eq. (4)] can be formed between two atomic orbitals on any adjacent atoms A and B. With p orbitals there are two possible orientations of the atomic orbitals. They can overlap end-on, as in Fig. 3a, or they can overlap sideways, as in Fig. 3b. In the former case the molecular orbital is designated as sigma (σ, "cylindrical"), and in the latter it is designated as pi (π) because the nodal plane of the atomic orbitals is preserved. Pi molecular orbitals must be used for the second bond of a double bond and for the second and third bonds of a triple bond.

E. Hybrid Orbitals

Placing two electrons into any of these molecular orbitals leads to stabilization since the electron probability is increased in the overlap region, where the electrons are attracted to two positive charges. However, because s and p atomic orbitals are symmetric, there is a limit to the extent to which that probability can increase. For any s or p orbital on atom A half of its probability is on the side opposite to atom B. To overcome this limitation, the s and p orbitals on an atom can be constructed into hybrid orbitals, defined as follows, where ψ_s, ψ_x, ψ_y, and ψ_z are respectively s, p_x, p_y, and p_z orbitals (Fig. 1), and the c's are coefficients:

$$\psi_{hybrid} = c_s\psi_s + c_x\psi_x + c_y\psi_y + c_z\psi_z. \quad (5)$$

Figure 4 shows that adding s and p_y orbitals leads to an increased probability of finding the electron to the right of the atom because in that region of space the s and p_y orbitals reinforce, whereas to the left the positive value of the s orbital tends to cancel the negative value of the p_y orbital and reduce the value of ψ_{hybrid} and also its square, which is the probability. The surface of constant ψ_{hybrid} is readily drawn, but usually a stylized version is adequate, as in Fig. 4.

These hybrids are then used to form the molecular orbitals, as in Eq. (4) but with a ψ_{hybrid} formed from atomic orbitals on atom A or B in place of ψ_A or ψ_B. If those hybrids point toward each other, as in Fig. 5, they can form a sigma bond of increased strength and stability because of an increased probability for the electrons to be found in the overlap region. The surface of constant ψ_{MO} is quite elaborate, and only the individual hybrid orbitals are shown, to suggest their boundary envelope.

Many hybrids are possible, depending on how the coefficients are chosen. However, the number of distinct hybrid orbitals must equal the number of atomic orbitals used to construct them. In a hybrid designated as sp^λ, the superscript λ represents the ratio of p character to s character in terms of the coefficients that enter Eq. (5):

$$\lambda = \left(c_x^2 + c_y^2 + c_z^2\right)^{1/2}/c_s. \quad (6)$$

The most common hybrids are sp^1 (customarily written simply sp), sp^2, and sp^3. No hybridization is possible to improve sideways overlap, so that one p orbital must be

FIGURE 2 Molecular orbital of the H_2 molecule, composed from the two individual atomic orbitals.

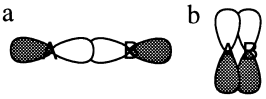


FIGURE 3 Two p orbitals forming (a) sigma and (b) pi molecular orbitals.

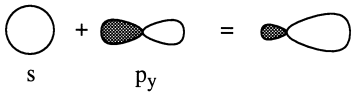


FIGURE 4 Hybrid atomic orbital formed from s and p_y orbitals.

FIGURE 5 Two hybrid orbitals forming a sigma molecular orbital.

dedicated to each pi bond. The remaining orbitals are hybridized to improve directionality. Then the hybridization of any (octet) atom is given by the following equation, in terms of the number of pi bonds or multiple bonds that the atom is forming:

$$\text{Hybridization} = \text{sp}^{(3-N_{\text{pi}})} = \text{sp}^{(3-N_{\text{double}}-2N_{\text{triple}})}. \qquad (7)$$

According to this formulation, the double bond of ethene (**4**) is composed of both sigma and pi molecular orbitals. The pi orbital is formed from sideways overlap of two p atomic orbitals on the carbons, as shown in Fig. 6a. According to Eq. (7), the remaining orbitals are of sp^2 hybridization. One of those hybrids on each carbon forms the C—C sigma bond, as shown in Fig. 6b. The other two sp^2 hybrids on each carbon, along with the s orbitals on the hydrogens, are used to form two C—H sigma bonds, shown in Fig. 6c.

Hybridization affects bond lengths and electronegativity. Because a p orbital has a node at the nucleus, an electron in that orbital has zero probability of being found at the nucleus, but there is no such restriction on an electron in an s orbital. Consequently, an s electron is closer to the nucleus and is held more tightly. Moreover, the distance of an electron from the nucleus and its ease of removal increase with the degree of p character. One manifestation of this feature is that C—H bond lengths decrease from 1.11 Å in ethane (CH_3CH_3, sp^3) to 1.10 Å in ethene ($CH_2{=}CH_2$, sp^2) to 1.08 Å in ethyne (HC≡CH, sp). Another is that the electronegativity of the carbon increases from sp^3 to sp^2 to sp.

The advantage of the orbital approach is that it provides a geometric picture. Since the sp^2 hybrid orbitals in ethene (Fig. 6) are constructed from p orbitals perpendicular to the one used for the pi bond, the molecule must be planar. However, for most purposes this approach is equivalent to the Lewis approach, where a molecular orbital [Eq. (4)] is symbolized simply by a line joining atoms A and B, regardless of whether it is sigma or pi.

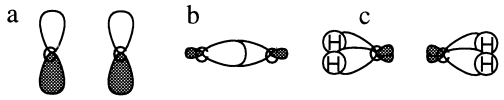


FIGURE 6 Molecular orbital description of ethene, $CH_2{=}CH_2$, viewed lying in a horizontal plane. (a) Atomic orbitals forming the pi molecular orbital. (b) Hybrid sp^2 orbitals forming the C—C sigma molecular orbital. (c) Hybrid sp^2 orbitals on carbon and s orbitals on hydrogens forming the C—H sigma molecular orbitals.

F. Antibonding Orbitals

When two atoms A and B interact, there are two solutions to Eq. (4), or two sets of coefficients c_A and c_B. One of those solutions is the bonding molecular orbital that allows electron density to accumulate between A and B. The other solution has a nodal surface between these atoms, and electrons are excluded from this region. This molecular orbital is said to be antibonding.

Usually such orbitals are empty, but this conclusion depends on the number of electrons available. If two helium atoms interact, there are a total of four electrons, only two of which are permitted in the bonding molecular orbital according to the Pauli exclusion principle. The other two must occupy the antibonding orbital. According to calculations, this orbital is more antibonding than the bonding orbital is bonding. The result is a net repulsion between the two helium atoms whenever they approach so closely that their orbitals overlap. Such repulsion between stable atoms or molecules is quite general. It is often called van der Waals repulsion or steric repulsion.

G. Molecular Orbital Theory

A unique advantage of molecular orbitals is that they provide an alternative to resonance theory for describing delocalized electrons. Equation (3) is generalized as follows, where $\psi_1, \psi_2, \ldots,$ and ψ_N are atomic orbitals on atoms 1 through N and the coefficients c_1–c_N are evaluated by quantum mechanical calculation:

$$\psi_{\text{MO}} = c_1\psi_1 + c_2\psi_2 + \cdots + c_N\psi_N. \qquad (8)$$

Usually only p orbitals are treated this way, and the sigma bonds are treated by one of the previous approaches. There are actually N different molecular orbitals, each with its set of coefficients. Some of these carry two electrons each and some are empty. For any of these the surface of constant ψ_{MO} is quite elaborate. Figure 7 suggests the boundary envelopes of two of the three filled pi molecular orbitals of benzene (**21**). Because these molecular orbitals extend over the entire molecule, they automatically allow for the electrons to be delocalized.

Calculations lead to the conclusion that cyclic systems containing $4n + 2$ ($n = 0, 1, 2, \ldots$) pi electrons are

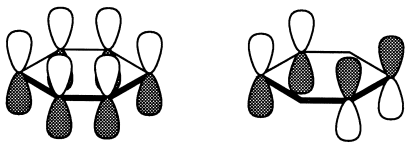


FIGURE 7 Two filled pi molecular orbitals of benzene, viewed lying in a horizontal plane and showing the p atomic orbitals. Hydrogens are omitted and carbons are not labeled.

unusually stable. Benzene (**21** or **24**; Fig. 7), with six pi electrons in the three pi bonds, is an example. The conclusion is very general, and other examples include cyclopentadienyl anion (**25**; $4n+2=6$), pyrrole (**26**; $4n+2=6$, including the lone pair), cyclopropenyl cation (**27**; $4n+2=2$), and naphthalene (**28**; $4n+2=10$ around periphery). In contrast, cyclic systems containing $4n$ ($n=1, 2, \ldots$) pi electrons, such as cyclobutadiene (**29**; $4n=4$) are unusually unstable. The unusual stability of cyclic systems containing $4n+2$ pi electrons is called aromaticity, and the instability of cyclic systems containing $4n$ pi electrons is called antiaromaticity. This is a distinction not accessible from considerations of resonance hybrids.

25 **26** **27** **28** **29**

Owing to the power of modern computers, it is now possible to calculate electronic structures, geometries, thermodynamic properties, and dynamic behavior of large molecules with a reliability of ±1 kcal/mole. The methods are quantum mechanical, of greater generality than Eq. (8). Software programs are widely available for performing such computations and for visualizing the results.

II. MOLECULAR STRUCTURE: STEREOCHEMISTRY

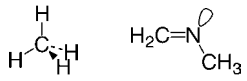
Stereochemistry is that aspect of molecular structure that deals with the geometry of molecules in three dimensions. Isomers are different molecules with the same atomic composition (same number of each kind of atom). For example, both CH_3OCH_3 and CH_3CH_2OH have two carbons, six hydrogens, and an oxygen but they are assembled differently. Likewise, $CH_3CH_2CH_2CH_3$ and $(CH_3)_3CH$ have the same atoms but different connectivity. Isomers that differ in connectivity are often called constitutional isomers. They have different Lewis structures. In contrast, stereoisomers have the same connectivity and the same Lewis structures but they are still different molecules.

Even the simplest aspects of stereochemistry often require a method for displaying the three-dimensional structure of molecules on a two-dimensional page. One such method is to use ordinary lines for bonds that are in the plane of the page, a wedge for a bond that comes forward, and a dotted line for a bond that goes behind the page. Another possibility is to draw a perspective view of the molecule, with thicker lines for bonds at the front (rather than shortening distant bonds, as in true perspective), or with bonds in front "concealing" bonds behind. Still another possibility is to sight along a single bond, show the bonds at the front atom as emanating from that atom, and show the bonds at the rear atom as partially hidden by the sigma bond, depicted as a circle. All of these methods are used below.

A. Bond Distances and Bond Angles

Bond distances are governed by the interplay between the accumulation of the electrons in the overlap region and the repulsions of the electrons for each other and of the nuclei for each other. For each bond there is an optimum distance that minimizes the total energy. That distance shortens with the number of electrons bonding the two atoms together. For example, the C—C distances in ethane (CH_3—CH_3), ethene (CH_2=CH_2), and ethyne (HC≡CH) are 1.54, 1.34, and 1.20 Å, respectively.

The optimum values for bond angles are governed largely by the mutual repulsion of electron pairs. For an octet atom it might be expected that its four electron pairs would be as far apart from each other as possible. The optimum arrangement then is to direct them toward the vertices of a regular tetrahedron. Methane (CH_4; **30**) is exactly such a molecule. The HCH angle is tetrahedral, 109.5°. Likewise, sp^3-hybridized carbons in other molecules have angles that are tetrahedral or nearly so. Angles at nitrogen and oxygen, as in NH_3 (**6**), H_2O (**7**), and related molecules, are also close to tetrahedral, to minimize the repulsions of the four electron pairs, regardless of whether they are in bonds or lone pairs. The deviations from the idealized 109.5° are due to a greater s character for lone pairs, which are held closer to the nucleus, or a greater p character for electrons that are unequally shared with a more electronegative atom.



30 **31**

Since every octet atom has four electron pairs, it might be thought that every such atom is tetrahedral. However, an unhybridized p orbital must be used for each pi bond and the remaining electron pairs then minimize their mutual repulsions. For an sp^2-hybridized atom, with three such pairs, the angle between them is 120°, and with two, on an sp-hybridized atom, it is 180°. Thus the HCH angle in ethene (CH_2=CH_2) and the CNC angle in CH_2=NCH_3 (**31**) are both close to 120° and the HCC angle in ethyne (HC≡CH) is 180°.

B. Conformations about Single Bonds

Conformations are different molecular structures related by rotation about single bonds. Since a sigma bond is cylindrically symmetric, there is little restriction to rotating around it. The restriction does not come from the

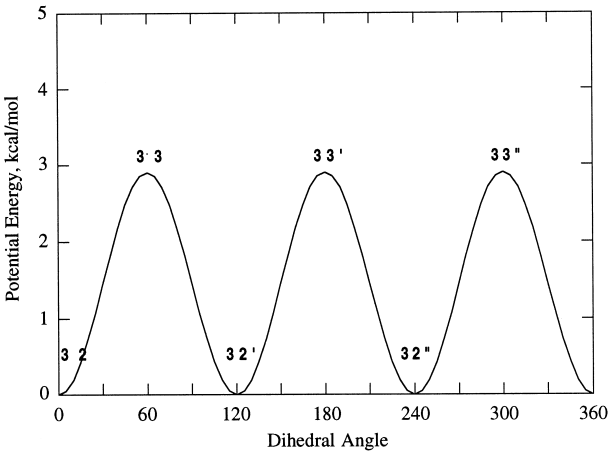


FIGURE 8 Energy of ethane versus dihedral angle (defined as zero for a staggered conformation).

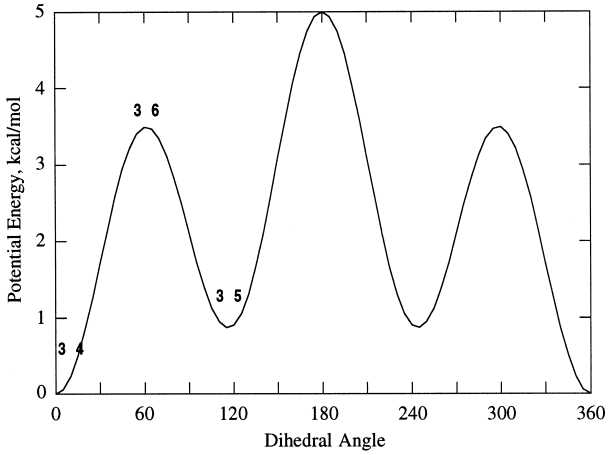
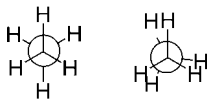


FIGURE 9 Energy of butane versus dihedral angle (defined as zero for the anti conformation).

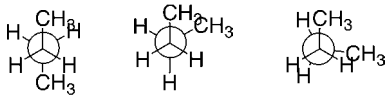
bond itself but from the interactions of the adjacent bonds with each other. In the simplest case, ethane, the most stable conformation is **32**. This is called the staggered conformation. The least stable is called the eclipsed conformation (**33**) because one hydrogen is directly behind the other (even though the drawing offsets them so that the rear hydrogens can be seen). The eclipsed conformation is 2.9 kcal/mole less stable than the staggered one. Figure 8 shows a graph of the conformational energy versus dihedral angle, the angle of rotation of one methyl group relative to the other. There are other staggered conformations (**32′** and **32″**) at 120° and 240° as well as other eclipsed conformations (**33′** and **33″**) at 180° and 300°.



32 **33**

Eclipsed conformations are destabilized by torsional strain. In ethane this arises from a subtle interaction of the electrons in one C—H bond with the electrons in the C—H bonds on the adjacent carbon. Those electrons repel, and the repulsion is minimized in the staggered conformation.

For butane there are two different kinds of staggered conformations, one called anti (**34**), with the two methyl groups as far apart as possible, and the other called gauche (**35**), with the methyls adjacent. Neither of these has torsional strain. The anti is most stable. The gauche is destabilized by steric repulsion, which arises because one of the hydrogens of one methyl is so close to one of the hydrogens of the other methyl that their orbitals overlap. This destabilization amounts to 0.9 kcal/mole. The conformational energy is graphed in Fig. 9. At a dihedral angle of 60° there is an eclipsed conformation (**36**) that is 3.4 kcal/mole less stable than the anti and 2.5 kcal/mole less stable than the gauche owing to torsional strain. At 180° there is another eclipsed conformation that is ~5 kcal/mole less stable than the anti owing to both torsional strain and steric repulsion.

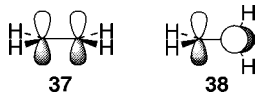


34 **35** **36**

A conformation that is a minimum on such an energy diagram is called a conformational isomer, or conformer. However, it is not customary to consider different conformers as isomeric or stereoisomeric because they interconvert very quickly. The energy required to rotate one group relative to another comes from the random thermal energy of molecular motion. If only 2.9 or 3.4 or 2.5 kcal/mole must accumulate, this happens very often (within picoseconds). Thus a chemical species is often present as a mixture of conformers. In some special cases conformers are interconverted "slowly" and these are called atropisomers.

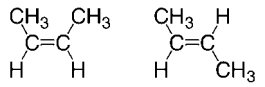
C. Isomerism about Double Bonds

In contrast to single bonds, rotation about double bonds is severely restricted. The second bond of a double bond is a pi bond, from sideways overlap of p orbitals on adjacent atoms. Figure 6 illustrates ethene, redrawn as **37**. To rotate one end of the double bond requires twisting one p orbital so that it does not overlap with the other, as in **38**. This structure is less stable by the energy of the pi bond, or ~60 kcal/mole. Random thermal energy cannot provide this much, so rotation around double bonds ordinarily does not occur.



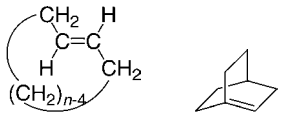
37 **38**

Consequently some molecules with double bonds can exist as a pair of stereoisomers, two different molecules with the same connectivity. For example, 2-butene exists as both (***Z***)-**39** and (***E***)-**39**. These designations rely on the Cahn–Ingold–Prelog priority rules. Older terminology is to call these *cis* and *trans*, respectively, and this sort of isomerism, arising from sidedness on a double bond, is often called *cis/trans* isomerism.



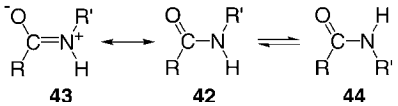
(***Z***)-**39** (***E***)-**39**

The (*Z*) isomer is 1.0 kcal/mole less stable than the (*E*) owing to steric repulsion between the two methyl groups. With cycloalkenes, such as cyclooctene and smaller rings, the (*Z*) or *cis* isomer is much more stable because bond angles are severely strained if a short chain must span the carbons on opposite sides of the double bond of the (*E*) isomer (**40**; $n = 5$–8). Likewise, bicyclic alkenes like **41** are severely strained by a double bond that can be (*Z*) or *cis* in one (six-membered) ring but must be (*E*) or *trans* in the other. The instability of such molecules is often called Bredt's rule.



40 **41**

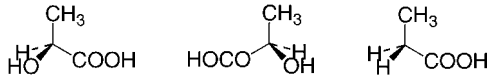
Amides (**42**) represent an intermediate situation. Because of the additional resonance form **43** there is partial C–N double-bond character. Rotating the NHR′ relative to the RC(=O) prevents the lone pair on the N from overlapping with the p orbital on carbon, similar to **38**. The twisted structure is less stable by the energy of the partial double bond, or ∼20 kcal/mole. Consequently there is another stereoisomer, **44**. The two stereoisomers interconvert rapidly at room temperature but they can be recognized as separate species by NMR techniques.



43 **42** **44**

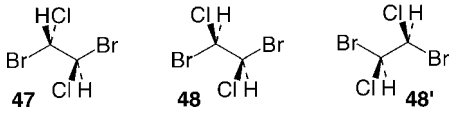
D. Stereoisomerism from Stereo Centers

A tetrahedral atom with four different groups attached is not superimposable upon its mirror image. Such an object is said to be chiral, and the atom is called a stereocenter. Lactic acid (2-hydroxypropionic acid, **45**, or its mirror image, **45**′) is such a molecule. No matter how **45**′ is rotated, it cannot be placed on top of **45** so that all atoms correspond. They have different configurations, or specific three-dimensional arrangements of the four groups around the central carbon. According to designations that rely on the Cahn–Ingold–Prelog priority rules, **45** is the (*S*) configuration and **45**′ is (*R*).



45 **45'** **46**

A molecule that is superimposable on its mirror image is said to be achiral. Propionic acid (**46**) is such a molecule. There are two simple structural features either of which guarantees achirality. One of these is a plane of symmetry through the middle of the molecule such that each atom not on that plane is matched by an identical atom reflected through that plane. The plane of the paper is a plane of symmetry for **46**. The other feature is an inversion center, a point in the middle of the molecule such that each atom is matched by an identical atom equidistant from that point and in the exact opposite direction. A molecule with this feature is (*R*, *S*)-1,2-dibromo-1,2-dichloroethane (**47**). This has two stereocenters but is nevertheless achiral. A molecule that is achiral despite having stereocenters is said to be meso.



47 **48** **48'**

A chiral molecule and its (nonsuperimposable) mirror image are two different molecules with the same connectivity. They are therefore stereoisomers. Such stereoisomers are said to be enantiomers. Enantiomers are very similar to each other, with the same melting point, boiling point, solubility, spectral characteristics, stability, and reactivity (except toward other chiral molecules). They differ in their ability to rotate the plane of polarized light, a property known as optical activity. One enantiomer will be dextrorotatory, rotating the plane of polarization to the right, whereas the other will be levorotatory, rotating the plane to the left. These are designated with prefixes (+)- and (−)-, respectively. The (*S*)-2-hydroxypropionic acid (**45**) is dextrorotatory (+)-2-hydroxypropionic acid. There is no universal relation between the *R* or *S* designation and the rotatory power.

Chirality is a necessity for optical activity, but it is not a guarantee. A 50:50 mixture of two enantiomers is optically inactive because the dextrorotatory and levorotatory powers exactly cancel. Such a material is called a racemic

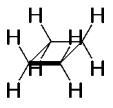
mixture, designated by a prefix (±)-. The process of separating the individual enantiomers from a racemic mixture is called resolution. The conversion of an optically active material into its racemic mixture is called racemization.

Diastereomers are stereoisomers that are not mirror images of each other. *Cis/trans* isomers like **39** are diastereomers. Another form of diastereomerism occurs with molecules that have two or more stereocenters. For example, a second stereoisomer of (*R*,*S*)-1,2-dibromo-1,2-dichloroethane (**47**) is (*R*,*R*)-1,2-dibromo-1,2-dichloroethane (**48**) and a third is (*S*,*S*)-1,2-dibromo-1,2-dichloroethane (**48′**). These last two are enantiomers of each other, but neither is the enantiomer of **47**. Each must then be a diastereomer of **47**. Even though diastereomers have identical connectivity and therefore all the same bonds, they are genuinely different from each other, with different melting points, boiling points, solubilities, spectral characteristics, stability, and reactivity.

This characteristic of diastereomers makes possible the resolution of a racemic mixture into its enantiomers. The process also takes advantage of the fact that chiral natural substances exist as single enantiomers. If the racemic mixture of (+) and (−) enantiomers can react with a natural substance that is (for definiteness) the (*R*) stereoisomer, the products are (+, *R*) and (−, *R*). These are diastereomers, with different solubilities, so they can be separated by treatment with a suitable solvent. Then the (+) enantiomer can be recovered from the (+, *R*) product and the (−) enantiomer from the (−, *R*).

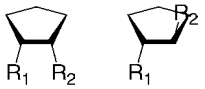
E. Cyclic Compounds

Cyclic compounds can show various combinations of these stereochemical features, including conformations, stereocenters, chirality, enantiomerism, diastereomerism, and another form of *cis/trans* stereoisomerism. The conformational behavior of cycloalkanes, C_nH_n, depends on the ring size. For $n = 3, 4$, or 5 the small ring constrains the CCC angle to $<109.5°$, leading to an instability called angle strain that arises from increased repulsion between electron pairs and poorer overlap. Also, there is torsional strain along each C—C bond, which is eclipsed in the planar conformation (e.g., **49**). For $n = 4$ or 5 small distortions from planarity reduce the torsional strain but at the expense of an increase of angle strain.



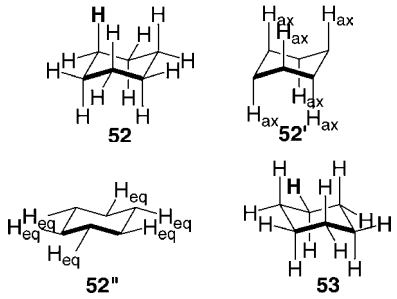
49

A disubstituted cycloalkane can have both of the substituents on the same face of the ring, as in **50**, or they can be on opposite faces, as in **51** (hydrogens omitted in both). These are diastereomers. The former is called *cis* and the latter is *trans*. If $R_1 \neq R_2$, both of these are chiral, so each can exist as a pair of enantiomers. The configuration at each stereocenter can be designated as (*R*) or (*S*). If $R_1 = R_2$, only **51** is chiral, since **50** has a vertical plane of symmetry between the two CHR carbons and is meso.



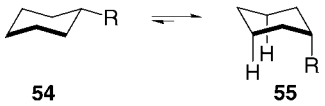
50 **51**

Cyclohexane can eliminate both angle strain and torsional strain by distorting from coplanarity to the chair conformation (**52**). This has the feature of axial and equatorial hydrogens in two distinctly different environments, as depicted in **52′** and **52″**, respectively. Rotation about the single bonds converts **52** into a different chair conformation, **53**, whereby each hydrogen that was axial (e.g., the boldface H at the rear of **52**) becomes equatorial (boldface in **53**) and each hydrogen that was equatorial becomes axial. This process, called ring inversion, occurs quickly (within microseconds at 25°C).



52 **52'** **52"** **53**

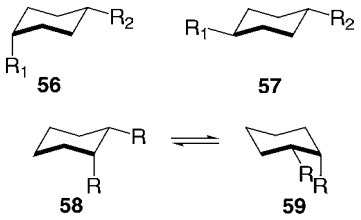
A monosubstituted cyclohexane thus exists as a rapidly equilibrating mixture of two conformers (**54**, **55**, most hydrogens omitted). The conformer with a bulky R substitutent in the axial position (**55**) is destabilized by steric repulsion with the two axial hydrogens nearby.



54 **55**

A disubstituted cyclohexane can be either *cis* or *trans*. If the substitution pattern is 1,4 (**56**, **57**), these are diastereomers. Both are achiral, as guaranteed by a vertical plane of symmetry through carbons 1 and 4. Each can undergo ring inversion, so each is a mixture of two conformers, only one of which is shown. If the substitution pattern is 1,2 or 1,3 and if $R_1 \neq R_2$, the *cis* and *trans* molecules are both chiral. If $R_1 = R_2$, the *cis*-1,3-disubstituted cyclohexane is achiral, but the *trans*-1,2, the *trans*-1,3, and the *cis*-1,2 (**58**) are all chiral. However, an unusual feature of this last is that ring inversion rapidly converts it to another

conformer (**59**) that is the enantiomer of **58**, so that this will always be present as a racemic mixture.



F. Molecular Mechanics

Each molecule has its own preferred geometry, determined by the various interatomic forces. If the preferences for optimum bond distances and bond angles and for minimum torsional strain and steric repulsion are characteristic of the atoms involved and independent of the rest of the molecule, then it becomes possible to extend our understanding from one set of molecules to others. It is necessary to express the dependence of energy on molecular geometry in a form like Eq. (9), where $V(\{x_i, y_i, z_i\})$ depends on the positions of all the individual atoms, and where V_{stretch} [Eq. (10)] represents the energy to distort the distance between atoms i and j from the optimum d_{ij}^0, V_{bend} [Eq. (11)] and V_{torsion} [Eq. (12)] represent the energy to distort the *ijk* bond angle or the *ijkl* dihedral angle from its optimum, and V_{es} and V_{vdW} are electrostatic and van der Waals (steric) energies, respectively, which depend on the distance between atoms i and j:

$$V(\{x_i, y_i, z_i\}) = \sum V_{\text{stretch}}(d_{ij}) + \sum V_{\text{bend}}(\theta_{ijk}) + \sum V_{\text{torsion}}(\theta_{ijkl}) + \sum V_{\text{es}}(d_{ij}) + \sum V_{\text{vdW}}(d_{ij}), \quad (9)$$

$$V_{\text{stretch}}(d_{ij}) = \tfrac{1}{2}k_{ij}\left(d_{ij} - d_{ij}^0\right)^2, \quad (10)$$

$$V_{\text{bend}}(\theta_{ijk}) = \tfrac{1}{2}k_{ijk}\left(\theta_{ijk} - \theta_{ijk}^0\right)^2, \quad (11)$$

$$V_{\text{torsion}}(\theta_{ijkl}) = \tfrac{1}{2}k_{ijkl}\left(\theta_{ijkl} - \theta_{ijkl}^0\right)^2. \quad (12)$$

The sums are over all bonds, angles, and atom pairs. The parameters that enter each of these terms can be calculated for model compounds by quantum mechanics. Alternatively they can be calibrated to give the best fit to an extensive set of experimental data on molecular geometries, stabilities, and infrared vibrational frequencies.

These calculations are formidable, but they are easy on a computer. Moreover, programs are available, complete with parametrization. Thus it is possible to calculate the energy for any arbitrary molecule in any geometry and to seek the geometry that minimizes that energy. It is further possible to compare that minimum with the energies of isomeric molecules and to calculate how the energy varies as the molecule is distorted to some other geometry.

III. CHEMICAL REACTIVITY

There are two aspects of chemical reactivity, static and dynamic. The static aspect relates to the question of chemical equilibrium, determined by the stability of reactants and products. The dynamic aspect relates to the question of chemical kinetics, or rates of reactions. This is more complicated, but it can be converted into static terms.

A. Chemical Equilibrium

For a general reaction, written as Eq. (13), the equilibrium constant is related to concentrations and to free energies by Eq. (14), where $\Delta G^\circ = G_{\text{B}}^\circ - G_{\text{A}}^\circ$, R is the gas constant, 8.314 J/mole or 0.001987 kcal/mole, and T is the absolute temperature:

$$\text{A} \rightleftharpoons \text{B}, \quad (13)$$

$$K = [\text{B}]/[\text{A}] = \exp(-\Delta G^\circ/RT). \quad (14)$$

Much of our understanding of chemical reactions comes from reasoning by analogy. If the equilibrium constant K is known for some standard reaction, it is often possible to predict, at least qualitatively, the equilibrium constant K' for a related reaction, involving some modification of the molecular structure. To do so, it is necessary to know how the modification affects energies.

For the modified reaction of Eq. (13′) if the modification stabilizes B′ (distinguished with a prime) relative to B, then it follows from Eq. (14) that $K' > K$ and that the equilibrium is shifted to the right:

$$\text{A}' \rightleftharpoons \text{B}'. \quad (13')$$

Likewise, if the modification destabilizes A, or raises its energy relative to A, then $K' > K$ and the equilibrium is again shifted toward the right. If the modification stabilizes A′ relative to A or destabilizes B′ relative to B, then $K' < K$ and the equilibrium is shifted toward the left. These conclusions are made graphic in Fig. 10.

Bond strengths represent a simple case of energetics that affect equilibrium. They are usually expressed as bond-dissociation energies (BDEs), positive numbers that correspond to the energy required to break a molecule into its constituent fragments, or the energy released when the bond forms. Table I lists some representative values.

Such values can be used to understand equilibria. For the reaction of Eq. (15) the total bond-dissociation energy of the C—H and Cl—Cl bonds on the left is 162 kcal/mole (104 + 58), whereas that on the right is 187 kcal/mole

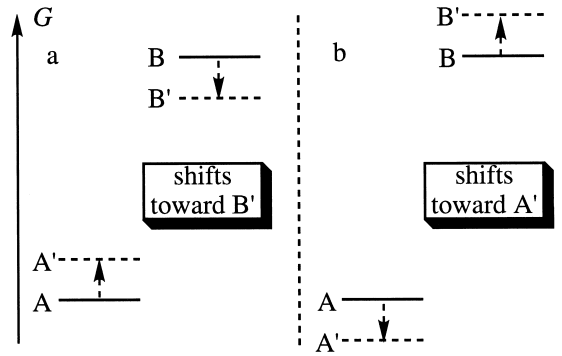


FIGURE 10 Effects of energy modifications on position of equilibrium A′ ⇌ B′ relative to A ⇌ B: (a) A′ destabilized or B′ stabilized (b) A′ stabilized or B′ destabilized.

(84 + 103). Therefore more energy is released when the products on the right are formed, meaning that this reaction is exothermic by 25 kcal/mole. In contrast, the corresponding reaction with iodine, Eq. (15′), is endothermic by 13 kcal/mole (104 + 36 − 56 − 71):

$$CH_3{-}H + Cl{-}Cl \rightleftharpoons CH_3{-}Cl + H{-}Cl, \qquad (15)$$

$$CH_3{-}H + I{-}I \rightleftharpoons CH_3{-}I + H{-}I. \qquad (15')$$

The modification of changing chlorine to iodine shifts the equilibrium to the left because the C—I and H—I bonds of the modified product are less stable than the C—Cl and H—Cl bonds of the unmodified product.

Other bond-dissociation data that are widely tabulated are acid-dissociation constants [Eq. (16), where HA is an acid in the reaction of Eq. (17)]. In contrast to the (homolytic) bond dissociations in Table I, these are heterolytic because when the H—A bond is broken, both electrons remain with A. Because acid-dissociation constants can range over many orders of magnitude, from the strongest acids, where $K_a \approx 10^{20}$, to the weakest, where $K_a \approx 10^{-50}$, the data are always tabulated in terms of pK_a [Eq. (18)]:

TABLE I Some Important A—B Bond-Dissociation Energies (kcal/mole)

	H	CH_3	OR	F	Cl	Br	I
H	104	104	110	135	103	87	71
CH_3	104	88	85	108	84	68	56
OR			35				
F				38			
Cl					58		
Br						46	
I							36

$$K_a = [H^+][A^-]/[HA], \qquad (16)$$

$$HA + H_2O \rightleftharpoons H_3O^+ + A^-, \qquad (17)$$

$$pK_a = -\log_{10} K_a. \qquad (18)$$

Acidity increases from left to right across the periodic table. For example, the pK_a values of CH_3—H, H_2N—H, HO—H, and F—H are ~50, 32.5, 15.7, and 3.2, respectively. This is a consequence of increasing electronegativity, which stabilizes the anions $F^- > OH^- > NH_2^- > CH_3^-$. Similarly, the acidity of hydrocarbons increases with carbon electronegativity from ethane (CH_3CH_3, sp^3) to ethene ($CH_2{=}CH_2$, sp^2) to ethyne (HC≡CH, sp). Of course electronegativity and acidity also increase with increasing positive charge, as for $H_2O < H_3O^+$ and $NH_3 < NH_4^+$.

B. Chemical Kinetics

Just because a reaction is exothermic or has a favorable equilibrium constant does not mean that it will occur. The reaction of Eq. (15) is exothermic by 25 kcal/mole and has a very large equilibrium constant, but if the reactants are mixed in the dark at room temperature, no reaction occurs. To understand such phenomena it is necessary to study chemical kinetics, the rates of chemical reactions.

For a general reaction [Eq. (19), slightly more elaborate than Eq. (13)] the rate of reaction can be defined by any of the forms of Eq. (20). The derivatives represent the decreasing concentration of reactants (hence the minus signs) or the increasing concentration of products as time passes. It is often found that the rate of reaction is proportional to some power (n_{AB} or n_C) of the concentrations of reactants (and perhaps to the concentrations of other species, such as catalysts), as in Eq. (21):

$$A{-}B + C \rightarrow A + B{-}C, \qquad (19)$$

$$\text{rate} = -\frac{d[AB]}{dt} = -\frac{d[C]}{dt} = \frac{d[A]}{dt} = \frac{d[BC]}{dt}, \qquad (20)$$

$$\text{rate} = k[A{-}B]^{n_{AB}}[C]^{n_C} \cdots. \qquad (21)$$

The constant of proportionality k is called the rate constant. It is a measure of how fast the reaction is.

C. The Transition State

In order to understand rate constants, it is necessary to specify a measure of the progress from reactants to products. This measure is a geometric parameter called the reaction coordinate. It is a composite of bond distances, chosen to increase in value from reactants to products. For

the reaction of Eq. (19) an appropriate reaction coordinate Q is given as follows since the distance between B and C is long in reactants and short in products, whereas the A—B distance increases as reaction proceeds:

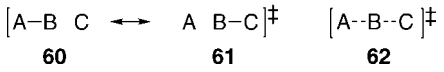
$$Q = d_{A-B} - d_{B-C}. \tag{22}$$

Therefore Q is a large negative number in the reactants and rises to a large positive number in the products.

The key to understanding reaction rates is the dependence of energy upon Q. As C approaches A—B, the energy must increase, from two sources: van der Waals repulsion arising from overlap of the electron clouds, and an endothermicity from starting to break the A—B bond. This energy continues to increase until eventually, at large Q, the energy will decrease as the B—C bond develops and as the van der Waals repulsion between B—C and A is relieved. Somewhere, in between, near $Q = 0$, when both bonds are stretched about equally, the energy reaches a maximum, as sketched in Fig. 11.

The position of the maximum is called the transition state, symbolized with a double dagger, ‡. This is the key concept for understanding chemical reactivity. It is a specific point along the reaction coordinate Q, or a unique geometry, with particular values of d_{A-B} and d_{B-C}. It is the geometry at which the energy is maximum.

The structure of the transition state cannot be described using simple Lewis structures since there are delocalized electrons. A pair of electrons is not localized in either the A—B bond, which is being broken, or in the B—C bond, which is being formed. The transition state can be described as a resonance hybrid of two resonance forms, **60** and **61**. Alternatively, in the shorthand form of **22**–**24** it can be symbolized as **62**, where the dotted lines symbolize the partial bonds that are breaking or forming. Yet according to the rules of resonance, only the electrons are delocalized. The transition state has its unique geometry, with nuclei fixed at the particular value of Q where the energy is maximum.



60 **61** **62**

That maximum energy of the transition state lies above that of the reactants by an amount $\Delta G^{\ddagger}$, equal to $G_{\ddagger} - G_{\text{reactants}}$, as shown in Fig. 11. This energy difference is called the activation energy or the free energy of activation. The rate constant of Eq. (21) can be related to this energy difference through the following equation, where k_B is Boltzmann's constant (1.381×10^{-23} J/K), h is Planck's constant (6.626×10^{-33} J-sec), R is the gas constant, and T is the absolute temperature:

$$k = (k_B T/h) \exp(-\Delta G^{\ddagger}/RT). \tag{23}$$

This equation arises because $\Delta G^{\ddagger}$ is the amount of energy that the reacting molecules must acquire in addition to that of the reactants. It is the barrier that must be overcome. This energy is used to push C into A—B and to break the A—B bond sufficiently for the B—C bond to start forming. That energy is derived from the accumulation of random thermal motions. The average thermal energy is RT, or about 0.6 kcal/mole at room temperature. Some molecules have less than average and some have more. Very few have as much as $\Delta G^{\ddagger}$, but those that do will react. Conversion of Eq. (23) into Table II shows how sensitive the rate of reaction is to both $\Delta G^{\ddagger}$ and to T. A low barrier corresponds to a reaction that is very fast. A reaction with a high barrier is very slow, but increasing the temperature makes it go faster.

Much of our understanding of chemical reactivity comes from reasoning by analogy. If the rate constant k is known for some standard reaction, it is often possible to predict, at least qualitatively, the rate constant k' for a related reaction involving some modification of molecular

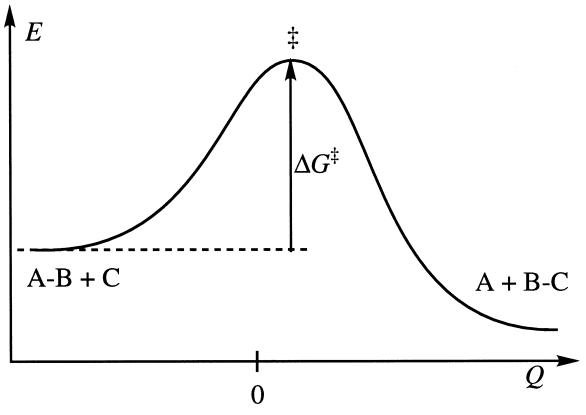


FIGURE 11 Variation of energy along reaction coordinate Q for the reaction A—B + C → A + B—C.

TABLE II Rate Constants and Activation Energy at Different Temperatures

$\Delta G^{\ddagger}$	At 25°		At 100°	
(kcal/mole)	k (sec^{-1})	Half-life	k (sec^{-1})	Half-life
0	6.21×10^{12}	0.112 psec	7.78×10^{12}	0.089 psec
5	1.34×10^{9}	516 psec	9.17×10^{9}	75.6 psec
10	2.91×10^{5}	2.39 μsec	1.08×10^{7}	64.1 nsec
15	6.28×10^{1}	11.0 msec	1.28×10^{4}	54.4 μsec
20	1.36×10^{-2}	51.0 sec	1.50×10^{1}	46.1 msec
25	2.94×10^{-6}	65.5 hr	1.77×10^{-2}	39.1 sec
30	6.35×10^{-10}	34.5 years	2.09×10^{-5}	9.21 hr
35	1.37×10^{-13}	160 kyears	2.47×10^{-8}	325 days
40	2.97×10^{-17}	739 Myears	2.91×10^{-11}	755 years

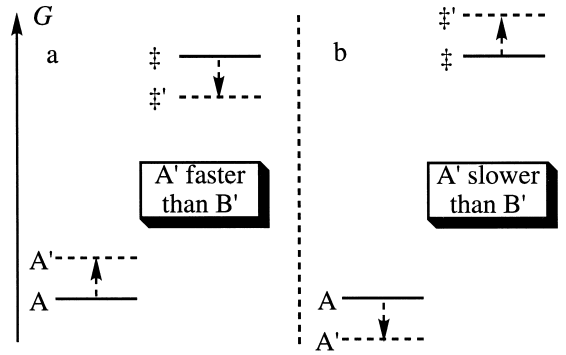


FIGURE 12 Effects of energy modifications on rate of reaction of A′ relative to A via transition state ‡ or ‡′: (a) A′ destabilized or ‡′ stabilized, (b) A′ stabilized or ‡′ destabilized.

structure. To do so, it is necessary to know how the modification affects energies, just as for equilibria above. For a general reaction written as follows, the rate constant is still given by Eq. (23):

$$\mathrm{A} \rightarrow \ddagger \rightarrow \mathrm{B}. \qquad (24)$$

For the following modified reaction (distinguished with a prime), if the modification stabilizes the transition state ‡′ (*not* the product B′), then it follows from Eq. (23) that $k' > k$ and that the reaction is faster:

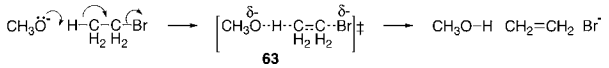
$$\mathrm{A}' \rightarrow \ddagger' \rightarrow \mathrm{B}'. \qquad (24')$$

Likewise, if the modification destabilizes the reactant A′ relative to A, the reaction is again faster. If the modification stabilizes reactant A′ or destabilizes transition state ‡′, then $k' < k$ and the reaction is slower. These conclusions are made graphic in Fig. 12, which is very similar to Fig. 10.

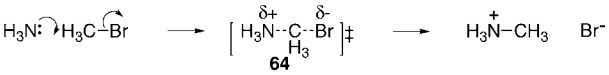
D. Electron Pushing

Because the transition state has an electronic structure describable as a resonance hybrid, electron pushing, which provided a convenient method for generating additional resonance forms, also provides a method for generating transition states. An electron pair is delocalized toward another atom so as to form a new bond, while also removing an electron pair from that adjacent atom if necessary to avoid violating the octet rule. The electron pair can come from a lone pair or from a multiple bond or a single bond. For example, this method permits the generation of the transition states for the methoxide-induced E2 elimination of HBr from ethyl bromide (**63**), the nucleophilic substitution of ammonia on methyl iodide (**64**), the electrophilic addition of H^+ (from HBr) to propene (**65**), the hydride transfer from borohydride ion to methanal (**66**), and the cycloaddition of ozone to ethene (**67**).

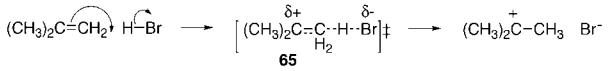
In such reactions the atom that carries the electron pair is called a nucleophile since it uses its electron pair to form a bond to some other nucleus. The partner with which the nucleophile reacts is called an electrophile since it forms a bond with that electron pair. In some contexts these are called Lewis bases and Lewis acids, respectively. One of the unifying features of physical organic chemistry is the recognition that a very large proportion of all reactions can be classified as ones where a nucleophile (or Lewis base) reacts with an electrophile (or Lewis acid). The vastness of the possible chemistry then arises from the great diversity of nucleophiles and electrophiles.



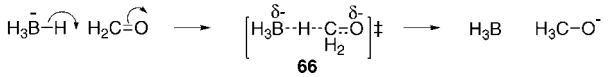
63



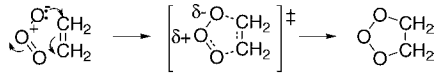
64



65

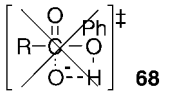


66



67

This method is also applicable to multistep reactions. For example, the hydrolysis of phenyl acetate, $CH_3C(=O)–OPh$, with hydroxide to form $CH_3CO_2^-$ plus PhOH does not proceed via transition state **68**, involving simultaneous breaking and making of the various bonds. Instead it is a three-step reaction, proceeding via three sequential transition states, **69**–**71**, each of which can be generated by electron pushing.



68

Between those transition states there are reaction intermediates **72** and **73**. These are ordinary chemical species, not transition states, but they are not very stable and they do not persist. Figure 13 shows how the energy varies during the course of the reaction, as measured by a reaction coordinate that is a composite of the various bond distances. The transition states are at local maxima and the intermediates are at local minima. In such a diagram there is generally one transition state that is higher than the others. That one is called THE transition state and its step is called the rate-limiting step. In the hydrolysis of phenyl

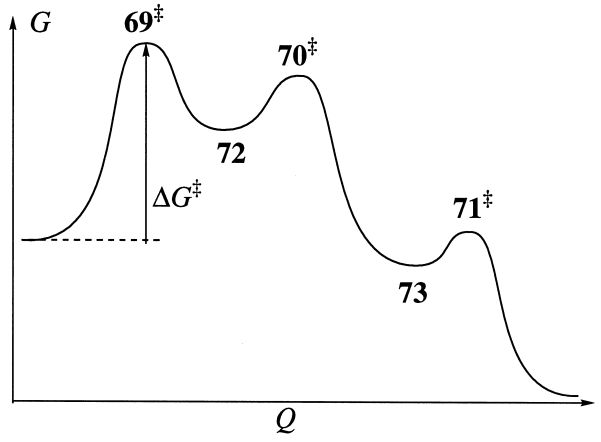
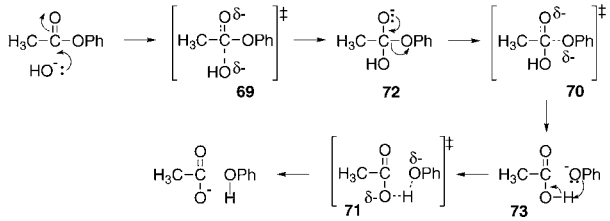


FIGURE 13 Variation of energy along reaction coordinate Q for the reaction $RC(=O)-OPh + OH^- \rightarrow RCO_2^- + PhOH$.

acetate it is the first transition state (**69**) that is highest, so that the first step is rate-limiting.



One of the principal goals of physical organic chemistry is to determine the mechanisms of reactions. A mechanism is a detailed description of the sequence of steps, involving breaking and making of bonds, that converts reactants into products, along with a description of the structure and energy of reaction intermediates and transition states along the way.

A notational system for representation of mechanisms has been devised. Bond making is indicated as A (association) and bond breaking as D (dissociation). A nucleophilic or nucleofugic (loss of a nucleophile) process at a core atom is indicated with a subscript N, or else E for electrophilic or electrofugic. A subscript H indicates hydrogen as electrophile or electrofuge at a core atom, but h if at a peripheral atom and xh if intermolecular. Simultaneous processes are indicated by juxtaposition, and successive steps are separated by a +. Thus the nucleophilic substitution corresponding to **64** is D_NA_N, the elimination corresponding to **63** is $A_{xh}D_HD_N$, and the sequence of **69**–**71** is $A_N + D_N + A_{xh}D_h$. In the older (Ingold) system these were designated as S_N2, E2, and $B_{Ac}2$, respectively.

E. Inductive Effects

One of the simplest modifications that can affect chemical reactivity is the introduction of a substituent that produces an electrostatic effect. The energy of interaction of two point charges q_1 and q_2 with each other or of one point charge with a dipole moment μ [Eq. (2)] is given as follows, where r is the distance between the two, ε is the dielectric constant of the solvent, and θ is the angle between the direction of the dipole and the direction toward the charge:

$$E_{\text{charge-charge}} = \frac{q_1 q_2}{\varepsilon r}, \tag{25}$$

$$E_{\text{charge-dipole}} = \frac{q\mu\cos\theta}{\varepsilon r^2}. \tag{26}$$

These correspond to an energy lowering when two opposite charges are near each other or when the charge is closer to the end of the dipole of opposite charge, and an energy increase when these are reversed.

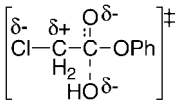
Table III lists acidity constants of some carboxylic acids RCOOH. Entry 1 is the reference acid, and the others differ because of electrostatic effects on the stability of the conjugate base, RCO_2^- (*not* the stability of the reactant). For entry 2 the negative charge on CO_2^- is stabilized by the positive charge on the NH_3^+ [Eq. (25)], so that the equilibrium of Eq. (17) shifts toward the right (decreased pK_a), according to the considerations in Fig. 10. For entry 3 the identical negative charges on the two CO_2^- groups destabilize each other [Eq. (25)], so that the equilibrium shifts toward the left (increased pK_a relative to entry 1). For entry 4 the positive end of the C—Cl dipole is closer to the negative charge on CO_2^- and leads to stabilization of the products [Eq. (26)] so that the equilibrium of Eq. (17) shifts toward the right, as for entry 2. Entry 5 is similar to entry 2 but the distance between the positive and negative charges is larger, so that the effect is smaller, according to the dependence of Eq. (25) on r.

These electrostatic effects are often referred to as inductive effects because they can also be viewed as arising from the substituents' pulling or pushing the electrons of the molecule and making it easier or harder for H^+ to be removed from the COOH. Inductive effects can also operate on rates, and they can be interpreted according to Fig. 12. For example, hydrolysis of $ClCH_2C(=O)OPh$ is faster than that of $CH_3C(=O)OPh$ because transition state

TABLE III Acidity Constants of Some Carboxylic Acids

Entry	Acid	pK_a
1	CH_3COOH	4.75
2	$^+H_3NCH_2COOH$	2.85
3	$^-O_2CCH_2COOH$	5.38
4	$ClCH_2COOH$	2.87
5	$^+H_3NCH_2CH_2COOH$	3.60

74 is stabilized relative to **69** by the interaction between the C—Cl dipole and the negative charge that is distributed over the O and the OH.



74

F. Solvent Effects

Solvent effects also arise from electrostatic interactions, but between solvent and solute. The energy of interaction of an ion or dipole is given approximately as follows, where q is the charge, μ is the dipole moment, r is its radius (the closest distance of approach between solvent and solute), and ε is the dielectric constant of the solvent:

$$E_{\text{solv,ion}} = \frac{q^2}{r}\frac{\varepsilon - 1}{2\varepsilon}, \tag{27}$$

$$E_{\text{solv,dipole}} = \frac{\mu^2}{r^3}\frac{\varepsilon - 1}{2\varepsilon + 1}. \tag{28}$$

The dielectric constant is an empirical measure of the polarity of a solvent, or of how well it stabilizes ions or dipoles. The second factor in each equation corresponds to a greater stabilization with increasing ε.

A polar solvent achieves that stabilization by clustering its own dipoles around the solute, as illustrated in Fig. 14. An anion or the negative end of a dipole is especially well stabilized by water because the H is so small that its $\delta+$ can approach quite close to the negative charge. The small r in Eqs. (27) and (28) then allows a large stabilization E. This is a general aspect of protic solvents, those with OH or NH groups. It is often attributed to hydrogen bonding, as though there were a bond between the H and the solute, but the interaction is largely electrostatic. In contrast, the $\delta+$ of polar aprotic solvents, such as dimethyl sulfoxide, $(CH_3)_2S{=}O$, is buried in the center of the molecule and cannot stabilize anions as well as the exposed $\delta-$ on the oxygen stabilizes cations.

Solvents have a large effect on reactions that create or destroy ions. For example, the acidity constant (Table III)

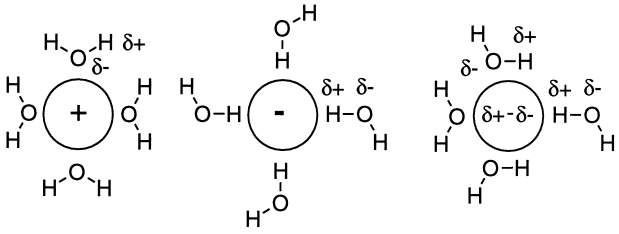


FIGURE 14 Solvation stabilization of solute cations, anions, or dipoles by water molecules.

of acetic acid is 4.75 in water, the standard solvent for pK_a, but is 9.65 in methanol. The dielectric constant of water is 81, but methanol is a less polar solvent, with a dielectric constant of only 33. Therefore, according to Eq. (27), methanol provides less stabilization of the ions H^+ and $CH_3CO_2^-$. In terms of Fig. 10, the change from water to methanol represents a relative destabilization of the products of ionization and therefore a decreased equilibrium constant (from $10^{-4.75}$ to $10^{-9.65}$).

Similarly, the solvolysis of *tert*-butyl chloride (SOH = H_2O or C_2H_5OH) is 3×10^5-fold faster in water than in ethanol as solvent:

$$(CH_3)_3C{-}Cl \rightarrow \left[(CH_3)_3C^{\delta+} \cdots Cl^{\delta-}\right]^{\ddagger} \rightarrow (CH_3)_3C^+ + Cl^- \tag{29}$$

$$(CH_3)_3C^+ + SOH \rightarrow (CH_3)_3CO(H)S^+ \rightarrow (CH_3)_3COS + H^+. \tag{30}$$

In this three-step reaction the first step is rate-limiting and its transition state has a large dipole moment, which can be stabilized by a polar solvent [Eq. (28)]. In terms of Fig. 12, the change from methanol to water represents a stabilization of the transition state and therefore a faster reaction.

Much current research involves the investigation of ions in the gas phase, free of the influence of solvation. A dominant influence is the size of the ion because a positive or negative charge constrained to a small volume repels itself strongly. Thus the order of gas-phase acidities of alcohols is $CH_3OH < CH_3CH_2OH < (CH_3)_2CHOH < (CH_3)_3COH$ because the negative charge is distributed over a larger volume in $(CH_3)_3CO^-$. This is exactly the opposite order from solution, where solvation of $(CH_3)_3CO^-$ is most hindered.

G. Delocalization Effects

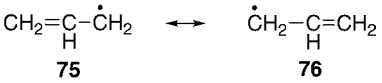
Delocalization of electrons generally leads to a stabilization. This is certainly true with both resonance and aromaticity, and the only exception is with antiaromaticity. Then, according to Fig. 10 or Fig. 12, delocalization can affect the position of equilibrium or the rate of reaction through stabilizing (or destabilizing) reactant, product, or transition state.

Table IV lists C—H bond-dissociation energies of some hydrocarbons. All the other bonds are weaker than the C—H of methane and require less energy to break. This weakening does not arise from differences in the hydrocarbons since all these C—H bonds are the same in that they are formed from an sp^3 orbital on carbon and an s orbital on hydrogen. Instead the weakening must come from stabilization of the radical that results from removing the

TABLE IV Some C—H Bond-Dissociation Energies (BDE)

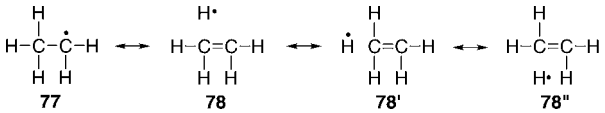
Entry	Bond	BDE (kcal/mole)
1	CH_3—H	104
2	CH_3CH_2—H	98
3	$(CH_3)_2CH$—H	95
4	$(CH_3)_3C$—H	92
5	CH_2=$CHCH_2$—H	89

hydrogen. For entry 5 that radical (**75**) is stabilized by an additional resonance form (**76**) which is not available for CH_3·. According to the rules of resonance, **76** is not just **75** flipped left-for-right, but it expresses the delocalization of the pi electrons in the double bond.



75 **76**

For the ethyl radical of entry 2 (**77**), resonance is less obvious, but three additional resonance forms are **78**, **78′**, and **78″**. According to the rules of resonance, these do not imply that hydrogens are removed. They symbolize only the delocalization of electrons from the adjacent C—H bonds to relieve the nonoctet character of the radical center at the expense of a hydrogen which loses its octet. In contrast to the delocalization of pi electrons in **75**–**76**, electrons in **77**–**78** are delocalized from C—H sigma bonds. This phenomenon is often called hyperconjugation. Similarly, for entries 3 and 4 there are hyperconjugative resonance forms that stabilize the radical that results from hydrogen removal. For the 2-propyl radical, $(CH_3)_2CH$·, of entry 3, there are a total of six additional forms, since electrons in each of the six adjacent C—H bonds can be delocalized. For entry 4 there are a total of nine additional forms. Thus the number of resonance forms for the radical increases and consequently the bond-dissociation energy decreases from entry 1 to 2 to 3 to 4.



77 **78** **78'** **78"**

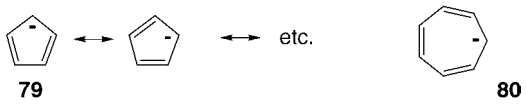
The heterolytic bond-dissociation energies of R—H to $R^+ + H^-$ (or of R—Cl to $R^+ + Cl^-$) also decrease along the series $R = CH_3 > CH_3CH_2 > (CH_3)_2CH > (CH_3)_3C$. Again this is due to hyperconjugation, which delocalizes electrons in the adjacent C—H bonds and relieves the nonoctet character of the carbocation at the expense of a hydrogen which loses its octet. In $CH_3CH_2^+$ there are three additional resonance forms, in $(CH_3)_2CH^+$ there are six, and in $(CH_3)_3C^+$ there are nine, paralleling the increased stabilization.

Similarly, the solvolysis of *tert*-butyl chloride is faster than that of isopropyl chloride. The transition state for the former is given in Eq. (29). That for the latter is given as follows:

$$(CH_3)_2CH\text{—}Cl \rightarrow [(CH_3)_2CH^{\delta+} \cdots Cl^{\delta-}]^{\ddagger} \rightarrow (CH_3)_2CH^+ + Cl^- \rightarrow \text{etc.} \quad (31)$$

However, those structures are incomplete. The transition state from isopropyl chloride has six additional hyperconjugative resonance forms, corresponding to delocalization of electrons in six adjacent C—H bonds. That from *tert*-butyl chloride has nine such forms, which confer greater stability. In terms of Fig. 12, the change from isopropyl to *tert*-butyl represents a stabilization of the transition state and therefore a faster reaction.

Another example is cyclopentadiene (pK_a 16), which is remarkably acidic for a hydrocarbon. This can be attributed to electron delocalization in the conjugate base (**79**, C's, H's, and lone pairs omitted), where there are five equivalent resonance forms, since the formal negative charge can be at any carbon. However, cycloheptatriene does not show any unusual acidity compared to other polyenes, even though its conjugate base (**80**) has seven equivalent resonance forms. This contrast is the consequence of the aromaticity of **79**, owing to the six pi electrons delocalized around the five-membered ring, whereas **80** is antiaromatic, with eight pi electrons.



79 **80**

H. Steric Effects

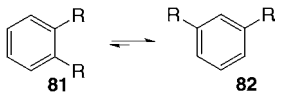
A steric effect arises from the sheer bulk of substituents. Just as two helium atoms repel, any two substituents repel whenever their orbitals overlap. In contrast to resonance, which is always stabilizing, steric repulsion is always destabilizing, although it can destabilize reactant, product, or transition state.

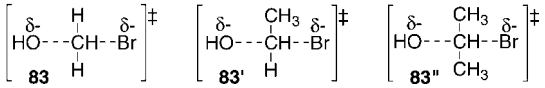
A simple example is the destabilization of *o*-xylene (**81**, $R = CH_3$) or *o*-di-*tert*-butylbenzene [**81**, $R = C(CH_3)_3$] by 0.3 and 22.3 kcal/mole, respectively, relative to their meta isomers (**82**). Another is the destabilization of axial conformer **55**. A quantitative measure of the effective bulk of a substituent *R* is the free-energy difference, $-RT \ln([\mathbf{55}]/[\mathbf{54}])$, often called the *A* value of the substituent. Some representative values are listed in Table V. These values can be used to estimate the position of the equilibrium between the two conformers of

TABLE V Energy Difference $A = -RT$ ln([55]/[54]) between Axial and Equatorial R

R	*A* (kcal/mole)
F	0.15
Cl	0.43
Br	0.38
CN	0.17
CH_3	1.74
CH_2CH_3	1.79
$CH(CH_3)_2$	2.21
$C(CH_3)_3$	>5.4
COOH	1.35
CO_2^-	1.92
OCH_3	0.60

a disubstituted cyclohexane like **56**. Still another example is given by the reactivity of alkyl bromides toward substitution by hydroxide, which decreases in the order $CH_3Br > CH_3CH_2Br \gg (CH_3)_2CHBr$. This is due to increasing destabilization of the transition state (**83**, **83′**, **83″**) by repulsion between OH or Br and zero, one, or two CH_3 groups on the central carbon.

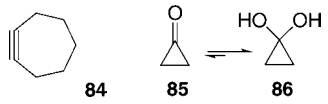




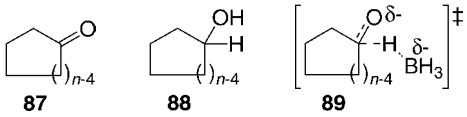
I. Strain Effects

Strain effects arise from the distortion of bond or torsional angles. The strain energy V can be expressed as in Eq. (11) or Eq. (12), where θ^0 is the optimum angle, of minimum energy.

Like steric effects, strain effects are always destabilizing, although they can affect reactant, product, or transition state. For example, cycloheptyne (**84**) is less stable than ordinary alkynes because the seven-membered ring prohibits the 180° angles preferred by the sp-hybridized carbons of the triple bond. Another example is the unusual behavior (compared to ordinary ketones) of cyclopropanone (**85**), which reacts with water to form cyclopropanediol (**86**). Both of these have strain energy because of the 60° bond angle in a three-membered ring. However, according to Eq. (11), the strain energy is greater for **85**, where the preferred bond angle of the sp^2 carbon is 120° ($V_{\text{bend}} = \frac{1}{2}k_{\text{CCC}}(60-120)^2 = 1800k_{\text{CCC}}$), compared to **86** ($V_{\text{bend}} = \frac{1}{2}k_{\text{CCC}}(60-109.5)^2 = 1225k_{\text{CCC}}$, which is smaller regardless of the magnitude of k_{CCC}).

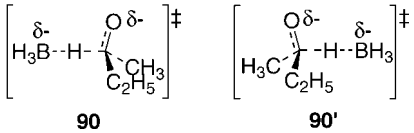


Torsional strain is responsible for the greater reactivity of cyclopentanone (**87**, $n = 5$) toward borohydride, as compared to that of cyclohexanone (**87**, $n = 6$), to form the alcohol (**88**, after protonation of the oxygen). The transition state (**89**, $n = 5$) leading to the five-membered ring is destabilized by increased torsional strain along all five C—C bonds relative to the reactant, where bonds to the carbonyl carbon are not eclipsed. Transition state **89** ($n = 6$) is stabilized by permitting all six C—C bonds to be staggered, eliminating any torsional strain. Therefore the transition state in the six-membered ring is more stable, accounting for the faster reaction.



J. Stereochemistry and Reactivity

How do stereoisomers differ in reactivity? The answer lies in the energetics of stereoisomeric transition states. Certainly enantiomeric transition states must have identical energies, since they are merely mirror images of each other. Therefore if two enantiomers react with an achiral reagent, they must react at identical rates. Moreover, if two enantiomeric products are formed from the same achiral reactant, they must be formed at equal rates and therefore in equal proportion. Only a racemic product can be formed. For example, the possible transition states for addition of borohydride to 2-butanone are enantiomers **90** and **90′**. The hydride is transferred to either of the two faces of the carbonyl group, but the two faces must be equally reactive.



In contrast, diastereomeric transition states have different energies. Therefore diastereomers show unequal reactivity and they can be formed at different rates. For example, addition of borohydride to (*R*)-3-methyl-2-pentanone, $CH_3C(=O)CH(CH_3)C_2H_5$, produces both (2*R*,3*R*)- and (2*S*,3*R*)-3-methyl-2-pentanol, but in unequal amounts. The reactant is chiral and there is an energetic preference for borohydride to be transferred to

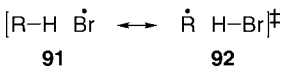
one of the two faces of the carbonyl. Much current research is devoted to understand the steric interactions that govern these preferences, with the aim of finding conditions to select one diastereomeric transition state over another.

Even if the reactant is (±)-3-methyl-2-pentanone, the products are equal amounts of (2*R*,3*R*)- and (2*S*,3*S*)-3-methyl-2-pentanol (a racemic mixture), plus equal amounts of (2*R*,3*S*)- and (2*S*,3*R*)-3-methyl-2-pentanol (another racemic mixture). However, the amount of the first pair is not equal to the amount of the second because the transition states are diastereomeric. The two transition states for formation of the first pair are enantiomeric, hence the racemic mixture. It is a general rule that only achiral products or racemic mixtures can be obtained from achiral or racemic reactants. It is not possible to obtain optically active product from optically inactive starting materials. The mechanism for the origin of the single enantiomers of chiral natural substances is not yet understood.

K. Relation between Kinetics and Thermodynamics

There is no universal relation between kinetics and thermodynamics. Just because a reaction is themodynamically very favorable does not mean that it must be fast. There are many reactions that are very exothermic but proceed extremely slowly. However, there is enough of a relation between kinetics and thermodynamics to make it useful for understanding.

Very often it is observed that in a pair of related reactions the one that is more favorable thermodynamically is the faster one. When a more stable product is formed because the transition state leading to it partakes of the stability of the product, this is said to be a case of product-development control. For example, the rates of reaction of bromine atoms with the hydrocarbons of Table IV follow the series $CH_3{-}H < CH_3CH_2{-}H < (CH_3)_2CH{-}H < (CH_3)_3C{-}H < CH_2{=}CHCH_2{-}H$. This order parallels the bond-dissociation energies, in that the weaker the bond is, the faster it cleaves. The transition state for any of these reactions can be written as a resonance hybrid of **91** and **92**, adapted from the generic **60** and **61**. However, there are additional resonance forms. Because **92** is a contributing resonance form, so are forms like **76** or like **78**, **78′**, and **78″**, which provide additional delocalization and additional stabilization. Consequently the transition states become increasingly stabilized and the reaction becomes faster.



91 **92**

It is cumbersome to consider transition states, with all their resonance forms and partial bonds. Instead it is easier to focus on products and recognize their similarity to the transition states. Then one can say that the more stable product is formed faster because it is more stable. This cannot be strictly true since rates depend on transition-state stabilities. However, if some feature stabilizes both transition state and product, then this statement is a convenient simplification. Thus the reactivity order of hydrocarbons toward bromine atoms may be attributed to the fact that radical stabilities increase in the order $CH_3\cdot < CH_3CH_2\cdot < (CH_3)_2CH\cdot < (CH_3)_3C\cdot < CH_2{=}CHCH_2\cdot$. Another example is the borohydride reduction of cyclopentanone (**87**, $n = 5$) and cyclohexanone (**88**, $n = 6$), where it is justifiable to say that the latter is faster because cyclohexanol is the more stable product owing to the absence of torsional strain.

In contrast, kinetic control is the case where one product is formed faster than another because of rate constants unrelated to equilibrium constants for the overall reaction. Such a case is the faster solvolysis of *tert*-butyl chloride [Eqs. (29)–(30)] compared to isopropyl chloride [Eq. (31)] even though the equilibrium constants for the two reactions are nearly the same. Therefore this is not product-development control. Yet it is possible to understand the greater reactivity of *tert*-butyl chloride because its transition state has nine additional hyperconjugative resonance forms. It is cumbersome to consider the transition states. Instead it is easier to focus on the carbocation intermediate, which is more stable in the *tert*-butyl case. Thus through an understanding of the mechanism, it is possible to understand this case of kinetic control in terms of the stability of intermediates that resemble the transition states.

IV. METHODOLOGY OF MECHANISTIC STUDIES

In the study of chemical kinetics one of the key pieces of information is how the reaction rate depends on reaction conditions. Among the variables are solvent, temperature, and the concentrations of the various reactants and catalysts and possibly other chemical species. Also, careful attention to product structures can provide additional information. However, it must be recognized that physical organic chemistry is an inductive science. It is never possible to prove a mechanism. At best it may be possible to obtain experimental results that are inconsistent with all conceivable mechanisms save one. Even then, there may be an "inconceivable" mechanism that was overlooked. Besides, for reasons of brevity experimental results are usually presented only as supporting a mechanism without

explicitly explaining how they are inconsistent with other mechanisms, and it is left to the critical reader to complete the logic.

A. Kinetic Order

It is often observed that v, the rate of a chemical reaction, is simply proportional to a power of the concentration of a chemical species:

$$v = -\frac{d[\text{reactant}]}{dt} = \frac{d[\text{product}]}{dt} = k[\text{A}]^{n_\text{A}}[\text{B}]^{n_\text{B}} \cdots . \tag{32}$$

Usually these species are reactants or perhaps catalysts, but they may be products or other additives. Such a reaction is said to be n_Ath order in A, n_Bth order in B, and $(n_\text{A} + n_\text{B} + \cdots)$th order overall. Usually reactions are first order, but some show second-order or zeroth-order dependence, and the exponent n can even be fractional or negative.

This is a differential equation that can be solved to express the time dependence of concentrations. It is an experimental task to determine each n. One way is to verify that the observed time dependence of concentrations follows that derived from solving Eq. (32). A better way is to vary the initial concentrations of the various chemical species and verify that v follows the power dependence of Eq. (32). For example, if a reaction is second order in a species, then v must quadruple if the concentration of that species is doubled. Another way is to hold all concentrations but one fixed and measure the dependence of v on that one variable concentration, using the following equation, derived from Eq. (32) by taking logarithms and partial derivatives:

$$\left(\frac{\partial \ln v}{\partial \ln[\text{A}]}\right)_{[\text{B}],\ldots} = n_\text{A}. \tag{33}$$

For example, the reaction of 2-bromopropane (RBr) with hydroxide, to form a mixture of 2-propanol and propene, shows a rate given by

$$\begin{aligned} v &= \{k_1 + k_2[\text{OH}^-]\}[\text{RBr}] \\ &= k_1[\text{RBr}] + k_2[\text{RBr}][\text{OH}^-]. \end{aligned} \tag{34}$$

This reaction is clearly first order in RBr. The order in hydroxide is undefined since it does not match the form of Eq. (32). However, the rate can be separated into two terms, one zeroth order in hydroxide and the other first order.

A convenient experimental technique is to follow the disappearance of one key reactant, called the substrate (S), while maintaining the concentrations of all other chemical species constant during each individual reaction. This can be accomplished (1) if the other species is in large excess, so that it is not consumed to any appreciable extent, or (2) if it is a catalyst, or (3) if it is H^+ or OH^- and the solution is buffered. The constancy of those other concentrations simplifies Eq. (32) to the form of Eq. (35) (or perhaps another form that is zeroth order or second order in substrate), where k_obs is called a rate coefficient (not a rate constant, because it varies with the concentrations of those other species). The solution is then Eq. (36), where S_0 is the initial concentration of substrate:

$$v = -\frac{d[\text{S}]}{dt} = k_\text{obs}[\text{S}], \tag{35}$$

$$[\text{S}] = S_0 \exp(-k_\text{obs} t). \tag{36}$$

The value of k_obs can thus be measured from the variation of [S] with time. By varying those other concentrations, the dependence of k_obs on those concentrations can then be evaluated experimentally.

In the example of Eq. (34) comparison with Eq. (35) shows that k_obs is given by

$$k_\text{obs} = k_1 + k_2[\text{OH}^-]. \tag{37}$$

By running the reaction in excess OH^- or in buffer, $[OH^-]$ can be kept constant during a reaction, to evaluate k_obs. By running the reaction with different concentrations of $[OH^-]$, the variation of k_obs with $[OH^-]$ can then be found to follow Eq. (37). The dependence of a k_obs on $[OH^-]$ or $[H^+]$ is often displayed as a pH–rate profile showing $\log_{10} k_\text{obs}$ versus pH. Figure 15 shows such a plot for the dependence of Eq. (37).

Kinetic orders give valuable information about mechanism. The kinetic order in a chemical species gives the number of those molecules in the transition state (strictly, the number of each of the constituent atoms that are required to form the transition state). For example, the first-order dependence on RBr in the reaction of Eq. (34) means

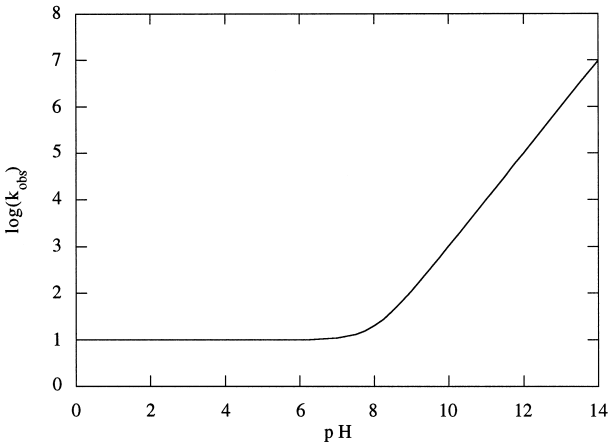


FIGURE 15 The pH–rate profile for a reaction where $k_\text{obs} = k_1 + k_2[\text{OH}^-]$.

that the atoms of one molecule of 2-bromopropane (three carbons, seven hydrogens, and a bromine) are present in the transition state. The dependence on $[OH^-]$ means that there are two pathways occurring together, one with an oxygen, an additional hydrogen, and an electron contained in the transition state, and the other without these.

One fundamental limitation of this method is that it is impossible to determine the kinetic order for the solvent molecules. It is not possible to vary the concentration of solvent without also introducing solvent effects of the sort arising via Eqs. (27) and (28). Therefore there is no way to know how many solvent molecules are present in the transition state, and the composition of the transition state is always subject to an uncertainty of an arbitrary number of solvent molecules. For example, the k_1 pathway in Eq. (34), which involves no hydroxide, might have a water molecule (two hydrogens and an oxygen) present in the transition state or it might not.

The first example of mechanistic inference from kinetics came from Lapworth's 1904 observation that the rate of bromination of acetone to form bromoacetone is first order in acetone and first order in acid but zero order in bromine:

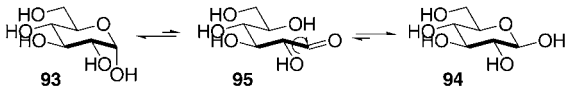
$$v = -\frac{d[CH_3COCH_3]}{dt} = \frac{d[CH_3COCH_2Br]}{dt} = k[CH_3COCH_3][H^+]. \quad (38)$$

Even though bromine is a reactant, the rate of reaction is independent of its concentration. These results mean that the transition state is composed of three carbons, seven hydrogens, an oxygen, and a positive charge (and an unknown number of the atoms that constitute water) but no bromines. Therefore the rate-limiting step occurs before the bromine enters the reaction. The current interpretation is that the rate-limiting step is proton removal by a water molecule from the conjugate acid of acetone to form the enol, $CH_3C(OH){=}CH_2$, as an intermediate that subsequently reacts rapidly with bromine.

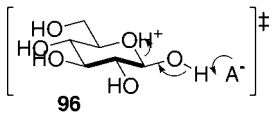
Another classic example is the nitration of benzene and other reactive aromatic hydrocarbons, where the rate depends on the concentration of nitric acid but is independent of the concentration of the aromatic. The current interpretation is that the rate-limiting step is formation of nitronium ion, NO_2^+, which then reacts rapidly with the aromatic.

The mutarotation of glucose is the interconversion of its α (**93**) and β (**94**) anomers, whose rate can be followed by the change of optical activity. The open-chain aldehyde (**95**) is an intermediate. The reaction is acid-catalyzed, but the rate is also found to increase with increasing buffer concentration, even at constant pH. Therefore the rate has the following form, where HA is the buffer acid:

$$v = k_{obs}[\text{glucose}] = \{k_H[H^+] + k_{HA}[HA]\}[\text{glucose}]. \quad (39)$$



Such behavior is called general acid catalysis since each acid is capable of serving as a catalyst. There must be one molecule of acid, either H^+ or buffer HA, present in the transition state, along with a molecule of the substrate. The current interpretation is that the transition state for reaction of **94** with HA has structure **96**, with H^+ detached from A^-. This example shows how the kinetic order tells us how many of each kind of atom are in the transition state but it does not itself tell us how those atoms are arranged.



B. Solvent Effects

Although it is not possible to determine the kinetic order in solvent, the fact that polar solvents can have large effects on reaction rates means that solvent effects can be used to diagnose whether a reaction creates or destroys ions. For example, reactions of trialkylsulfonium ions $RS(CH_3)_2^+$ with hydroxide show different solvent effects, depending on R. For $R = CH_3$ the rate decreases 2×10^4-fold on changing from ethanol to water, whereas for $R = (CH_3)_3C$ the decrease is only 3-fold. The large effect in the former case is consistent with a transition state $[HO^{\delta-} \cdots CH_3 \cdots S(CH_3)_2^{\delta+}]^{\ddagger}$, where ions are being destroyed. According to Fig. 12, this is a case where the more polar solvent, water, stabilizes the reactants and renders them less reactive. The small effect in the latter case is consistent with a transition state $[(CH_3)_3C^{\delta+} \cdots S(CH_3)_2^{\delta+}]^{\ddagger}$ that is still an ion because hydroxide is not yet involved.

There are empirical measures more suitable than dielectric constant for assessing quantitatively the polarity of a solvent. Even these do not always account for the specific interactions whereby a solvent stabilizes ions or dipoles. Table VI lists some relative rate constants for nucleophilic substitution of Cl^- on CH_3I (by the D_NA_N mechanism). There is no correlation with dielectric constant. The lower reactivity in the first three solvents is because they are protic, with $OH^{\delta+}$ or $NH^{\delta+}$ groups that approach close to chloride anion and stabilize it greatly ("hydrogen bonding"). To make the chloride react, it must be stripped of its solvation. In contrast, the last two solvents are aprotic, with their $\delta+$ buried in the center of the molecule, where it cannot stabilize anions as well. Consequently the

TABLE VI Relative Rate Constants for Reaction of Cl^- with CH_3I

Solvent	ε	k/k_{CH_3OH}
CH_3OH	32.6	1
$HC(=O)NH_2$	109.5	12.5
$HC(=O)NHCH_3$	165.5	45.3
$HC(=O)N(CH_3)_2$	37	1.2×10^6
$CH_3C(=O)N(CH_3)_2$	37.8	7.4×10^6

chloride is present as a "naked" nucleophile, of enhanced reactivity.

C. Temperature Dependence of Reactivity

Table II showed how rates can depend strongly on temperature. Since $\Delta G = \Delta H - T\Delta S$ the free energy of activation in Eq. (23) can be separated into entropy and enthalpy components, leading to the logarithmic form

$$\ln\left(\frac{k}{T}\right) = \ln\left(\frac{k_B}{h}\right) + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{R}\frac{1}{T}. \qquad (40)$$

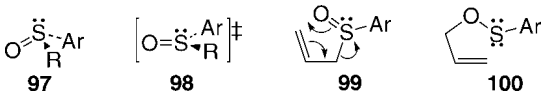
Therefore a plot of $\ln(k/T)$ versus $1/T$ has a slope equal to $-\Delta H^{\ddagger}/R$ and an intercept equal to $\ln(k_B/h) + \Delta S^{\ddagger}/R$, or $\Delta S^{\ddagger}/R + 23.76$. Thus both $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ can be evaluated experimentally.

These quantities are often called activation parameters. The enthalpy contribution generally comes from the energy required to break or reorganize the bonds. The entropy contribution generally comes from the need to organize the atoms into the precise arrangement of the transition state. Thus these parameters provide information about the nature of the transition state.

Table VII lists activation parameters for the racemization of some sulfoxides, RS(=O)Ar (**97**, $Ar = C_6H_4CH_3$-*p*). The sulfur is a tetrahedral stereocenter, owing to its four substituent groups (R, O, Ar, and lone pair). Racemization usually occurs by distorting the sulfur to an achiral transition state **98**, with R, O, and Ar coplanar. For entry 1 there is a substantial enthalpy of activation [Eq. (11)] because the sulfur must be distorted from tetrahedral bond angles to trigonal. There is no entropy contribution because reactant and transition state are equally well organized. For entry 2 racemization proceeds instead by fragmenting the molecule into $C_6H_5CH_2$ and ArSO radicals. There is a substantial positive entropy because the transition state is disorganized, owing to the creation of two particles from one. There is also a higher enthalpy than in entry 1 because a carbon–sulfur bond must be broken. For entry 3 racemization proceeds via transition state **99** and reversible conversion to an achiral intermediate **100**. There is a substantial negative entropy because the transition state is highly ordered, with the position of the $CH_2{=}CHCH_2$ group restricted. There is also a lower enthalpy than in entry 2 because formation of a carbon–oxygen bond compensates for breaking the carbon–sulfur bond.

TABLE VII Activation Parameters for Racemization of Sulfoxides, $RS(=O)C_6H_4CH_3$-*p*

Entry	R	$\Delta H^{\ddagger}$ (kcal/mole)	$\Delta S^{\ddagger}$ (cal/mole K)
1	C_6H_5	39	0
2	$C_6H_5CH_2$	43	24
3	$CH_2{=}CHCH_2$	22	−9



D. Isotope Effects

According to the Heisenberg uncertainty principle, it is impossible to determine exactly both the position and momentum of a particle. Therefore a hydrogen atom cannot be motionless at the distance r^0_{CH}, where the energy of a C—H bond (Fig. 16) is minimum, since then its position would be known, and also its kinetic energy and hence its momentum would be known to be zero. Instead it must have nonzero kinetic and potential energy, the sum of which is its zero-point energy. According to quantum mechanics, that energy is given as follows, where k_F is the ratio $F/(r^0_{CH} - r_{CH})$ of restoring force to distortion and μ is the reduced mass, which is approximately equal to m_H or m_D:

$$E_0 = \frac{1}{4\pi}\left(\frac{k_F}{\mu}\right)^{1/2}. \qquad (41)$$

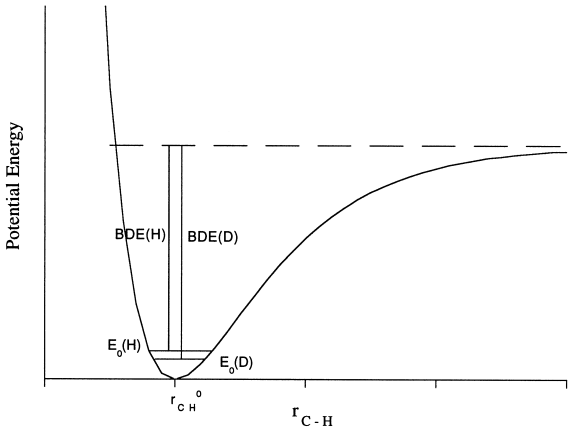
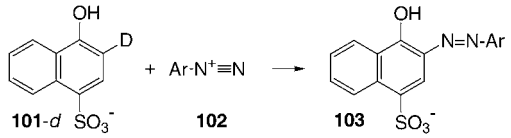


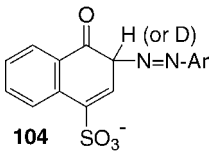
FIGURE 16 Energy to dissociate a C—H or C—D bond, each with its zero-point energy.

Because $m_D > m_H$ the zero-point energy is greater for a C—H bond than for a C—D, as shown in Fig. 16, and the bond-dissociation energy is less for C—H than for C—D. Then, according to Fig. 12, a C—H bond will break faster. For typical values of k_F, the rate constant ratio, k_H/k_D, is ~7 at 25°C.

This result can be used to determine whether a C—H bond is broken in the rate-limiting step. The comparison can be either intermolecular, between deuterated and undeuterated reactants, or intramolecular, by analyzing deuterium content in products from a partially deuterated reactant. For example, part of the evidence for transition states like **63** is the observation that $CH_3CHPhCH_2Br$ reacts with $NaOC_2H_5$ 7.5 times as fast as does $CH_3CDPhCH_2Br$. In contrast, the reaction of 1-naphthol-4-sulfonate-2-*d* (**101**-*d*) with *o*-methoxybenzenediazonium ion (**102**, Ar = *o*-$CH_3OC_6H_4$) to form **103** proceeds at the same rate as that of the undeuterated **101**. Therefore the C—H or C—D bond is not broken in the rate-limiting step, which is inferred to be the formation of **104** as an intermediate that loses H or D in a subsequent step.



101-*d* 102 103



104

The applicability of isotope effects is still wider. They can be seen even when a C—H or C—D bond is not broken but when the carbon bearing H or D undergoes rehybridization, although the effect is considerably smaller. They can be seen when reaction is carried out in D_2O rather than water. They can be detected with other isotopes, such as of carbon, nitrogen, or oxygen. Again the effects are smaller.

E. Regiochemical and Stereochemical Studies

It is a truism that the products of a reaction must be known in order to have any success at proposing a mechanism. Moreover, careful attention to product structures can provide important information about mechanism.

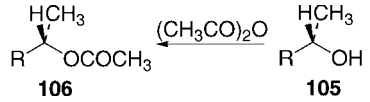
A classic puzzle was the mechanism of addition of HBr to alkenes such as propene. Sometimes the product was 2-bromopropane and sometimes it was 1-bromopropane. Eventually the role of the reaction conditions was recognized. The former arises by electrophilic addition via transition state **65**, leading to the carbocation intermediate $(CH_3)_2CH^+$ stabilized by six hyperconjugative resonance forms. The alternative regiochemistry would proceed via the intermediate $CH_3CH_2CH_2^+$, with only three such forms (from delocalization of electrons in two C—H bonds and one C—C). However, in the presence of peroxides, bromine atoms are generated and a free-radical chain process intrudes [Eq. (42) followed by Eq. (43)]:

$$Br^{\cdot} + CH_3CH{=}CH_2 \rightarrow CH_3CH^{\cdot}{-}CH_2Br, \quad (42)$$

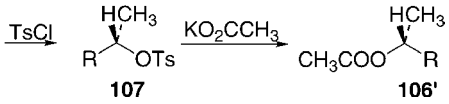
$$CH_3CH^{\cdot}{-}CH_2Br + HBr \rightarrow Br^{\cdot} + CH_3CH_2CH_2Br. \quad (43)$$

The regiochemistry here is due to the six hyperconjugative resonance forms in the intermediate $CH_3CH^{\cdot}$—CH_2Br, which is thus more stable than the alternative, $CH_3CHBrCH_2^{\cdot}$, which has only three hyperconjugative forms (from delocalization of electrons in one C—H bond, one C—C, and one C—Br).

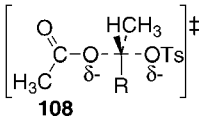
Many nucleophilic substitutions proceed with inversion of configuration at the carbon undergoing substitution. One of the first demonstrations of this behavior was with (−)-2-octanol (**105**, R = C_6H_{13}), which could be converted with acetic anhydride to (−)-2-octyl acetate (**106**). Alternatively, **105** could be activated by conversion to its tosylate (**107**, Ts = $SO_2C_6H_4CH_3$-*p*). Reaction with potassium acetate then formed (+)-2-octyl acetate (**106′**), the enantiomer of **106**. Conversion of levorotatory to dextrorotatory did not occur in the reactions with acetic anhydride (**105** → **106**) or tosyl chloride (**105** → **107**), which react at oxygen, not carbon. Therefore in the reaction of **107**, the nucleophile ($CH_3CO_2^-$) and the leaving group (TsO^-) must be on opposite faces of the carbon undergoing substitution. Transition state **108** shows this geometry (with the dotted C—C bond behind the plane of the page and with the other two dotted lines symbolizing partial bonds). It should be noted that this conclusion and the next were reached without knowing the absolute configurations of the molecules.



106 105



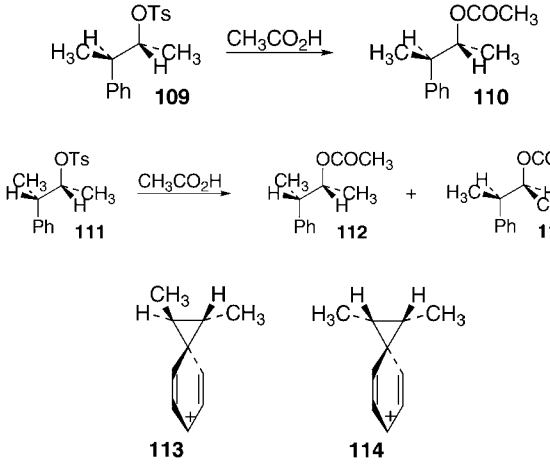
107 106'



108

In contrast, (2*R*,3*R*)-3-phenyl-2-butyl tosylate (**109**, Ts = $SO_2C_6H_4CH_3$-*p*) reacts in acetic acid to form

(2*R*,3*R*)-3-phenyl-2-butyl acetate (**110**), with retention of configuration. Even more remarkable is that (2*R*,3*S*)-3-phenyl-2-butyl tosylate (**111**) reacts in acetic acid to form equal amounts of (2*R*,3*S*)- and (2*S*,3*R*)-3-phenyl-2-butyl acetate (**112** + **112′**). Again the reaction proceeds with retention of configuration, but here the product is racemic. These are reaction conditions that would favor as rate-limiting step the formation of a carbocation, $PhCH(CH_3)CH^+CH_3$, rather than attack by acetate, as with **107**. However, that carbocation would be the same from either **109** or **111**, which should then have given the same products. Instead, to avoid so unstable an intermediate, the pi electrons of the phenyl group serve as an internal nucleophile. As in **108**, the configuration at C2 (the carbon originally attached to oxygen) is inverted as phenyl substitutes for tosylate. The resulting intermediate is a phenonium ion, **113** (from **109**) or **114** (from **111**). When this then reacts with acetic acid as nucleophile, the configuration at C2 is again inverted. Two successive inversions amount to an overall retention of configuration, as observed from both **109** and **111**. However, **114** is achiral, with a plane of symmetry passing through the six carbons of the benzene ring. Therefore the product must be racemic. This phenomenon of an internal nucleophile acting to avoid formation of an unstable carbocation is often called neighboring-group participation.



F. Labeling and Crossover Experiments

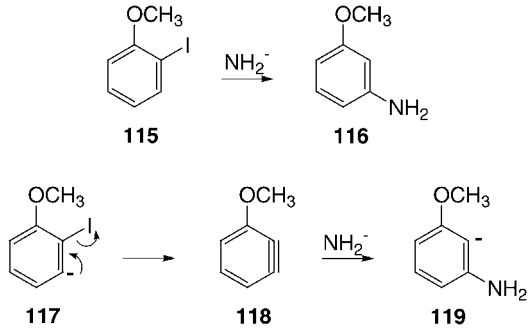
Labeling experiments are a means to tag a portion of a molecule and follow it through the reaction. The label may be an isotope or it may be a chemical substituent, which opens the risk of changing the mechanism but which is often easier to synthesize and to analyze.

Isotopic labeling shows that ester hydrolysis generally proceeds with cleavage of the acyl–oxygen bond, not the alkyl–oxygen bond. For example, hydrolysis of *n*-pentyl acetate in alkaline $H_2{}^{18}O$ produces acetate containing the ^{18}O and *n*-pentanol without any ^{18}O:

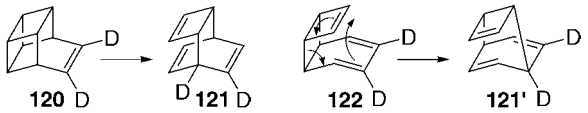
$$CH_3C(=O)OC_5H_{11} + {}^{18}OH^- \rightarrow CH_3C(=O){}^{18}O^- + C_5H_{11}OH. \qquad (44)$$

This experiment could also be performed with $CH_3C(=O){}^{18}OC_5H_{11}$ to verify that the $^{18}O{-}C_{pentyl}$ bond does not cleave. However, it is easier to "label" that position with ordinary ^{16}O and run the reaction in $H_2{}^{18}O$. A rare exception to this behavior is with the highly hindered methyl 2,4,6-tri-*tert*-butylbenzoate, where hydrolysis in $H_2{}^{18}O$ produces $CH_3{}^{18}OH$ by alkyl–oxygen cleavage.

Reaction of aromatic halides with $NaNH_2$ in liquid ammonia produces the corresponding aromatic amine. However, the conversion of *o*-iodoanisole (**115**) to *m*-anisidine (**116**) shows that this is not simply a substitution of NH_2 for I. Instead it proceeds by proton removal to create **117**, which undergoes elimination to a benzyne (**118**) that preferentially adds NH_2^- at the meta position to produce **119**. The methoxy group is a label to make the rearrangement clear (and to stabilize the anion in **119**), and the mechanism was further documented through ^{14}C labeling.

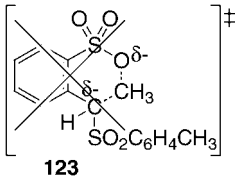


Another example is the rearrangement of **120** to **121**. The deuterium labeling shows that the reaction does not proceed simply by opening the four-membered ring at the left. Instead it proceeds by opening the vertical bonds in the middle to form **122**, followed by rearrangement as indicated (to **121′**, identical with **121**).



Crossover experiments are a form of double-labeling experiment designed to distinguish between intramolecular and intermolecular mechanisms. For example, methyl transfer from oxygen to carbon in *o*-(p-$CH_3C_6H_4SO_2$)$CH^-C_6H_4SO_2OCH_3$ to form *o*-(p-$CH_3C_6H_4SO_2$)$CH(CH_3)C_6H_4SO_3^-$ was shown to be intermolecular by using a mixture of this anion (d_0) plus (p-$CD_3C_6H_4SO_2$)

$CH^-C_6H_4SO_2OCD_3$ (d_6) and observing that a statistical mixture of d_0, d_3, and d_6 products was formed. The intramolecular methyl transfer would require transition state **123** but with a 180° CH · · · CH_3 · · · O angle, to be consistent with inversion of configuration at the methyl, and this geometry is impossible. This reasoning is known as the endocyclic restriction test.



123

G. Reactive Intermediates

In multistep reactions there are reaction intermediates between the transition states. Intermediates are ordinary chemical species at local minima in the energy, but they are not very stable and they are so reactive that they do not persist. Therefore they are at such low concentrations and of such short lifetimes that it is often impossible to detect them in the reaction mixture. Evidence for intermediates sometimes comes from the kinetic behavior. For example, an enol intermediate was inferred from the observation that the rate of bromination of acetone [Eq. (38)] is independent of $[Br_2]$. Another example is the absence of a kinetic isotope effect in the reaction of **101**-*d* with **102**, showing that the C—H or C—D bond is not broken in the rate-limiting step, which is the formation of **104** as an intermediate.

A simple kinetic model for a reaction with one intermediate is Eq. (45). The implied differential equations can be simplified with the steady-state approximation. If the concentration of the intermediate B does not build up appreciably, then its time derivative can be neglected relative to other variations, as expressed in Eq. (46). This algebraic equation is readily solved to give Eq. (47). This then leads to a simpler differential equation for [A] that has the form of Eq. (35) with k_{obs} given by Eq. (48):

$$\mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{B} \xrightarrow{k_2} \mathrm{C}, \tag{45}$$

$$\frac{d[\mathrm{B}]}{dt} = k_1[\mathrm{A}] - k_{-1}[\mathrm{B}] - k_2[\mathrm{B}] \sim 0, \tag{46}$$

$$[\mathrm{B}] = \frac{k_1}{k_{-1} + k_2}[\mathrm{A}], \tag{47}$$

$$k_{obs} = \frac{k_1 k_2}{k_{-1} + k_2}. \tag{48}$$

If $k_{-1} \ll k_2$, this reduces to k_1 and the first step is rate-limiting. If $k_{-1} \gg k_2$, this reduces to $(k_1/k_{-1})k_2$ and the second step is rate-limiting. The magnitude of k_1 is irrelevant to the question of which step is rate-limiting, but if $k_1 \ll k_{-1} + k_2$, then the concentration of intermediate B must be very small.

In some cases k_{-1} or k_2 in Eqs. (45)–(48) is not a rate constant but a rate coefficient, depending on the concentration of some other species. If so, it may be possible to vary that concentration so that the first step is rate-limiting under some conditions and the second step is rate-limiting under others. The change of rate-limiting step then appears as a characteristic dependence of rate on that concentration. For example, Fig. 17 shows the pH–rate profile for the reaction of acetone with excess hydroxylamine to produce acetoxime, $(CH_3)_2C{=}NOH$. The falloff at low pH is due simply to the conversion of hydroxylamine to unreactive $HONH_3^+$. The falloff at high pH is due to a change of rate-limiting step from formation of $(CH_3)_2C(OH)NHOH$ as intermediate at low pH to acid-catalyzed dehydration of that intermediate at high pH, where this step becomes slow.

The existence of intermediates can also be inferred from other kinds of evidence. The retention of configuration in reaction of **109** and the racemization in reaction of **111** are evidence for intermediate phenonium ions (**113** and **114**). The rearrangement in the reaction of **115** with $NaNH_2$ suggests a benzyne intermediate (**118**). The variable regiochemistry in the addition of HBr to alkenes was taken as evidence for both the carbocation $(CH_3)_2CH^+$ and the radical $CH_3CH^{\cdot}{-}CH_2Br$, depending on conditions.

Trapping experiments are often used to test for intermediates. If the chemical reactivity of a presumed intermediate can be predicted, then addition of a suitable reactant may intercept the intermediate and convert it to a distinctive product. A further control experiment is necessary to verify that the disappearance of substrate is not accelerated

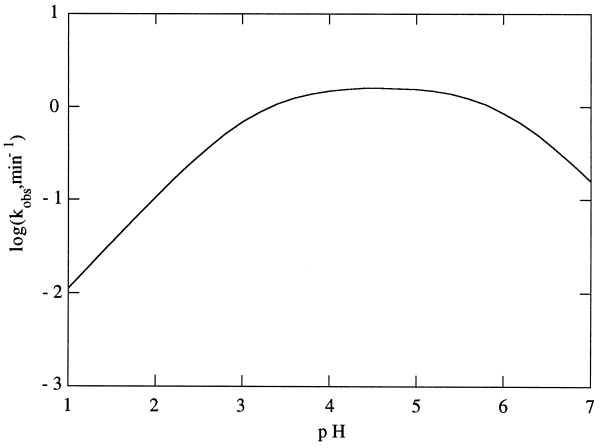


FIGURE 17 The pH–rate profile for reaction of acetone with hydroxylamine.

by the addition, in order to rule out direct reaction between additive and substrate. For example, the hydrolysis of $HCCl_3$ in base to HCO_2^- (plus CO) is thought to proceed via CCl_3^- and the carbene CCl_2. Indeed, if the hydrolysis is performed in D_2O and unreacted chloroform reisolated, it is found to be largely $DCCl_3$, from trapping of CCl_3^-. Moreover, adding PhS^- does not increase the rate of disappearance of $HCCl_3$, but the product is $HC(SPh)_3$ instead, from trapping of CCl_2. Another example is the solvolysis of Ph_2CHCl or Ph_2CHBr, both of which are thought to proceed via Ph_2CH^+ as a common intermediate. Added sodium azide (NaN_3) does not increase the rate (except through its effect on solvent polarity) but does change the product from Ph_2CHOH to Ph_2CHN_3.

A variant on trapping experiments is to test substances that would inhibit a free radical reaction. For example, certain phenols, like BHA and BHT in food products, can scavenge free radicals. If the reaction does not take its normal course, then it is not necessary to detect a distinctive product in order to conclude that free radical intermediates are implicated.

A convincing procedure is to observe the intermediate. This is particularly informative since it permits the intermediate to be characterized spectroscopically and its properties elucidated. Usually the intermediate is too unstable to be observed under the reaction conditions. One possibility is to modify the substrate so as to stabilize the intermediate and increase its lifetime. Alternatively, it may be possible to modify the conditions so that the intermediate persists. For example, treatment of triphenylmethanol with sulfuric acid converts it to the carbocation Ph_3C^+, stabilized by resonance delocalization of the positive charge into the three aromatic rings. Less stable carbocations can be prepared in "superacid" solutions of $HSbF_6$, where there is no nucleophile available to react, and where they can be characterized by NMR spectroscopy. Similarly, many intermediates can be preserved by eliminating other species that they might react with. Thus carbanions can be observed under conditions that scrupulously exclude water and other electrophiles. For example, the basicity of enolates and related anions can be measured by titration in dimethyl sulfoxide, and Table VIII lists some acidity constants of their conjugate acids. Preserving free radicals and carbenes (R_2C) is more challenging since they react with themselves, but the remedy is to immobilize them in a solid matrix so that they cannot encounter another reactive partner. Reactive intermediates that rearrange by themselves, such as strained structures (**40** with $n = 6$–8, **41**, **84**) and some carbenes and carbocations, can often be generated at low temperature, perhaps photochemically. Similarly, electronically excited states can be generated photochemically and their behavior studied.

TABLE VIII Acidity Constants of Some Carbonyl, Nitro, Cyano, and Sulfone Compounds in Dimethyl Sulfoxide

Acid	pK_a
$CH_3COCH_2COOC_2H_5$	14.2
$CH_2(COOC_2H_5)_2$	16.4
CH_3NO_2	17.2
$PhCOCH_2Ph$	17.65
$PhCH_2CN$	21.9
$CH_3COCH_2SO_2Ph$	22.1
$PhCH_2COOC_2H_5$	22.6
$(PhCH_2)_2SO_2$	23.9
$PhCOCH_3$	24.7
Cyclopentanone	25.8
CH_3COCH_3	26.5
$[(CH_3)_2CH]_2CO$	28.2

A variant is independent synthesis of a presumed intermediate, which offers the further opportunity to test whether it is converted to product (or, even better, to the identical mixture of products) under the reaction conditions. For example, it is possible to prepare salts of NO_2^+ and to show that these react with aromatic hydrocarbons to form the same nitration products as with nitric acid.

Another possibility is that the presumption of intermediacy must be rejected. For example, the Clemmensen reduction of ketones $R_2C{=}O$ to hydrocarbons R_2CH_2 with zinc metal had been thought to proceed via the alcohol R_2CHOH. However, this cannot be an intermediate because it is found to be inert to the reaction conditions.

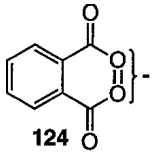
V. ILLUSTRATIVE EXAMPLES

A. Computation and Structure

The location of the hydrogen in a hydrogen bond is an intriguing question of molecular structure that can now be addressed confidently by high-level computations. Usually a hydrogen is bonded covalently to one oxygen and hydrogen-bonded to the other. If the two oxygens are of equal or nearly equal basicity, the hydrogen may be bonded to either and jump quickly from one to the other [Eq. (49)]. In contrast, for the monoanion of phthalic acid (**124**), recent quantum mechanical calculations indicate a symmetric structure, with the hydrogen centered between the hydrogens and bonded equally to both. This conclusion is quite sensitive to the oxygen–oxygen distance $d_{O–O}$, which must be short. With a highly accurate method known as density functional theory $d_{O–O}$ for the

gas-phase ion was computed to be 2.38 Å. It does not change if the ion is embedded in a continuum whose dielectric constant is equal to that of chloroform or water. An alternative approach to solvation is to simulate the motions of the ion and of discrete solvent molecules, as governed by their forces of interaction. Interaction with chloroform molecules does not change d_{O-O}, but if the ion also interacts with K^+, it becomes asymmetric. In water d_{O-O} increases to 2.46 Å and the structure again becomes asymmetric because of the entropy associated with hydrogen bonding by water to the ion. These results are in agreement with some experimental data.

$$O{-}H\cdots O \rightleftharpoons O\cdots H{-}O. \qquad (49)$$

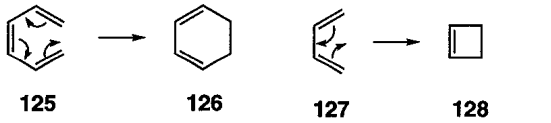


124

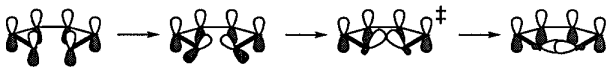
B. Pericyclic Reactions

The Woodward–Hoffmann rules are a triumph of the application of molecular orbital theory to reactivity. They are concerned with pericyclic reactions, reactions where electrons reorganize around a cycle of nuclei, with simultaneous bond breaking and bond making, as in **67**. There are three classes of pericyclic reactions: (1) electrocyclic reactions, where a new single bond is made across the ends of a pi system, as well as the reverse of such a reaction, (2) sigmatropic rearrangements, where a group migrates across a pi system, and (3) cycloaddition reactions, where one pi system adds across another, as well as the reverse, a cycloreversion reaction. The full treatment of the Woodward–Hoffmann rules involves consideration of the symmetry of the molecular orbitals, but there is a simpler approach that that focuses on aromaticity of transition states.

The electrocyclizations of 1,3,5-hexatriene (**125**) to 1,3-cyclohexadiene (**126**) and of 1,3-butadiene (**127**) to cyclobutene (**128**) offer a revealing comparison. From the initially planar triene (**125**), rotation about the outermost pi bonds allows overlap between the p atomic orbitals on carbons 1 and 6, to form **129**. If those carbons also rehybridize toward sp^3, the transition state (**130**) is reached. Eventually those carbons form a sigma bond and the product (**126**) results. Similarly, the planar butadiene (**127**) rotates and rehybridizes to create transition state **131**, which proceeds to form the sigma bond of the product (**128**). Unlike the acyclic reactants or the products, where the sigma bonds do not overlap with the pi orbitals, the transition states may be characterized by a cycle of atomic orbitals, each overlapping with two neighbors. Electrons are delocalized over those cycles since they are in bonds that are forming and breaking. In **130** there are six delocalized electrons, and in **131** there are four. Therefore transition state **130** is aromatic and **131** is antiaromatic. These reflect stabilization or destabilization universally associated with cyclic systems containing $4n+2$ ($n=0, 1, 2, \ldots$) or $4n$ ($n=1, 2, \ldots$) delocalized electrons. Accordingly, **130** represents a low-energy transition state, corresponding to a fast reaction, whereas **131** is of high energy and corresponds to a slow reaction. The conclusion that **125** can undergo this reaction whereas **127** cannot is a distinction not accessible from resonance theory. It does follow from molecular orbital theory but it requires a generalization to transition states of the concept of aromaticity.



125 **126** **127** **128**



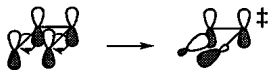
125 **129** **130** **126**

There are two alternatives whereby **127** can react. To account for these it is necessary to generalize aromaticity even further, to excited states and to a different topology. According to molecular orbital calculations, the excited electronic states of cyclic systems containing $4n$ delocalized electrons are stabilized (relative to acyclic comparisons), whereas those with $4n+2$ are destabilized. Thus aromaticity and antiaromaticity reverse in an excited state. For example, the excited electronic state of transition state **131** becomes stabilized, and **127** can be transformed rapidly into **128** under photochemical conditions.

127 **131** **128**

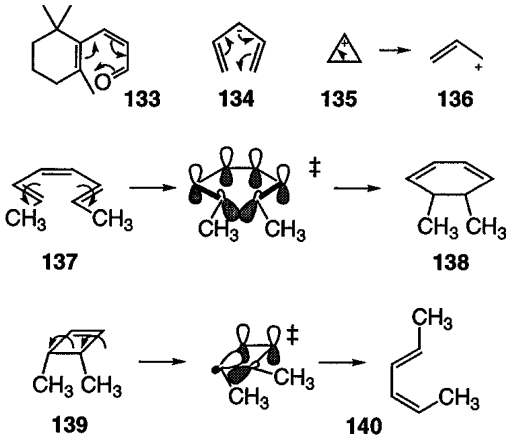
The alternative topology arises by rotating the outermost pi bonds of **127** in the same direction. This is called conrotatory motion, in distinction to the previous disrotatory motion, where the bonds rotated in opposite directions. This now leads to transition state **132**, where the top lobe of the rehybridizing orbital on carbon 1 overlaps with the bottom lobe of the orbital on carbon 4. It would seem as though this is an antibonding interaction because a nodal surface is created between these atomic orbitals. However, according to molecular orbital calculations, this leads to a stabilization. Indeed, this is a general result for cyclic systems containing $4n$ delocalized electrons but subject to

the requirement for overlap between the positive lobe of one orbital and the negative lobe of another. Such an arrangement of orbitals has the topology of a Möbius strip, obtained from a rectangular ribbon by twisting one end by 180° before fastening it to the other end. The normal topology, often called Hückel in this context, always has positive lobes overlapping positive (or negative overlapping negative). According to molecular orbital calculations, stabilization and aromaticity can be associated with all Möbius cycles containing $4n$ delocalized electrons, and destabilization and antiaromaticity with $4n + 2$.



127 **132**

Thus, thermally activated electrocyclizations, or the reverse reactions, occur in disrotatory fashion when there are $4n + 2$ pi electrons, and conrotatory when there are $4n$. It also follows that these conclusions must be reversed for photochemical reactions. These are very general conclusions, applicable also to molecules with heteratoms (**133**), to anions (**134**), and to the ring opening of **135** to **136**. However, the topology becomes apparent only with labels. For example, disrotatory closure of *trans,cis,trans*-2,4,6-octatriene (**137**) produces *cis*-5,6-dimethyl-1,3-cyclohexadiene (**138**), and conrotatory opening of *cis*-3,4-dimethylcyclobutene (**139**) produces *cis,trans*-1,3-butadiene (**140**), even though the *trans* and *trans,trans* stereoisomers of the products would be favored by stability.

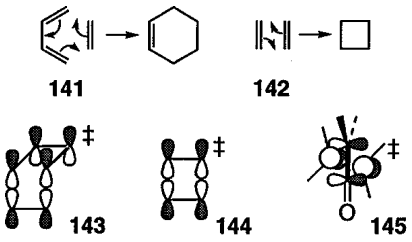


133 **134** **135** **136**

137 **138**

139 **140**

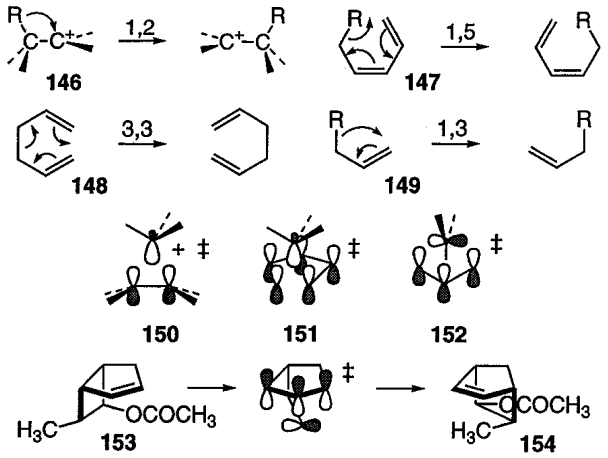
In cycloadditions one pi system adds across another, as in the Diels–Alder reaction (**141**) of butadiene plus ethylene to produce cyclohexene (or a wide variety of substituted versions), or the dimerization of ethene (**142**) to cyclobutane. These proceed via transitions states **143** and **144**. The former is aromatic, with six electrons delocalized over the cycle of atomic orbitals. The latter is antiaromatic, with four electrons delocalized. Therefore the former reaction is quite facile, but the latter does not occur thermally. It does occur photochemically, with appropriate substituents to facilitate creation of the excited electronic state. Alternatively, a cycloaddition with $4n$ delocalized electrons can occur if the orbital overlaps can achieve a Möbius topology. This requires addition across opposite faces of a pi systems. Such an addition is called antarafacial, in contrast to the usual suprafacial, across a single face of a pi system. Usually this latter is sterically the only possibility. However, an alkene can add across a ketene $R_2C{=}C{=}O$ via transition state **145**, with overlap between the positive lobe of an alkene carbon and the negative lobe of the carbonyl carbon.



141 **142**

143 **144** **145**

Sigmatropic rearrangements are the most bewildering of pericyclic reactions because there is always a reacting pair of electrons from a sigma bond. This feature can be avoided in the other cases by focusing on the electrocyclization or cycloaddition direction, even when the reverse reaction is under consideration. Examples are the 1,2 alkyl shift of a carbocation (**146**), the 1,5 alkyl shift of a diene (**147**), the 3,3 shift of two three-atom fragments across each other (**148**), and the 1,3 alkyl shift of an alkene (**149**). The transition states for the first two (**150**, **151**) have $4n + 2$ (two or six) electrons delocalized over a cycle of atoms. Similarly, the 3,3 shift has six delocalized electrons in its transition state. All of these are facile reactions, with low activation energies, because they proceed via aromatic transition states. The 1,2 alkyl (or hydride) shift (**146**) is a remarkably fast process that is responsible for many rearrangements that carbocations undergo. In contrast, the 1,3 alkyl shift (**149**) has four delocalized electrons in its transition state. The suprafacial process therefore involves a high-energy, antiaromatic transition state and is slow. However, if the geometry permits an antarafacial process, with migration of the three-atom component across opposite faces of the single atom, then there is one overlap between a positive lobe and a negative lobe. This is a Möbius topology and thus an aromatic transition state (**152**). A remarkable example of the success of the Woodward–Hoffmann rules is the rearrangement of **153** to **154**, with what would otherwise have been a surprising change of the relative stereochemistry of methyl and acetoxy.



These rules are summarized in Table IX. To judge whether a reaction can occur via a low-energy aromatic transition state, all that is necessary is to find the cycle of overlapping atomic orbitals in that transition state, to ascertain whether all the overlaps can be between lobes of the same sign or whether there must be one positive-to-negative overlap, and then to count the number of electrons delocalized over that cycle.

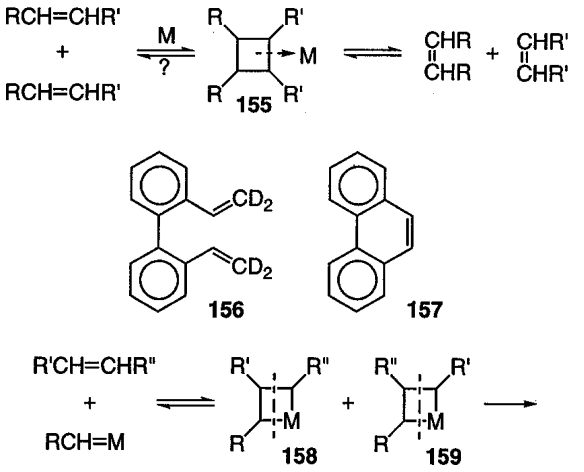
C. Organometallic Catalysis

Olefin metathesis is the interchange of fragments on double bonds, as in Eq. (50). The reaction is catalyzed by various compounds of tungsten, molybdenum, rhenium, or ruthenium. The mechanism had been thought to involve some sort of complex (**155**) between two alkenes and a metal (with additional, unspecified ligands, and whose d orbitals may permit a transition state of Möbius topology). However, when the reaction was carried out with a mixture of 2,2′-divinylbiphenyl and 2,2′-divinylbiphenyl-d_4 (**156**), the product, even at short reaction times, was phenanthrene (**157**) plus a statistical mixture of ethene-d_0, -d_2, and -d_4. If an intermediate like **155** were formed from either of these reactants, it could not cleave to ethene-d_2. Therefore it was proposed that the catalytically active form of the metal is a metallocarbene, RCH═M, which reacts with an alkene R′CH═CHR″ to form both **158** and **159** as intermediates. These can cleave to RCH═CHR′ and RCH═CHR″ and also regenerate catalytic metallocarbenes R′CH═M and R″CH═M.

TABLE IX General Rules for Pericyclic Reactions

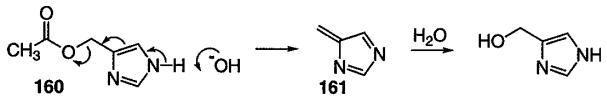
Conditions	$4n+2$ electrons	$4n$ electrons
Thermal	Hückel	Möbius
Photochemical	Möbius	Hückel

$$CH_3CH{=}CHC_2H_5 \rightleftharpoons CH_3CH{=}CHCH_3 + C_2H_5CH{=}CHC_2H_5. \quad (50)$$



D. Ester Hydrolysis

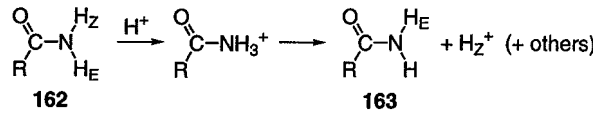
In alkali, 4-(acetoxymethyl)imidazole (**160**) is hydrolyzed $\sim 10^5$ times as fast as benzyl acetate, CH_3COOCH_2Ph. The customary mechanism of ester hydrolysis, via addition of hydroxide to the carbonyl carbon (**69**), should not show such a rate disparity. Therefore a different mechanism, not available to benzyl acetate, was proposed, proceeding via intermediate **161**. As evidence, when the hydrolysis was carried out in alkaline $H_2{}^{18}O$, the isotope was found in the alcohol product, 4-(hydroxymethyl)imidazole, whereas ester hydrolysis ordinarily transfers the isotope to the acetate [Eq. (44)]. Moreover, when the NH was replaced by NCH_3, no rate enhancement was observed.

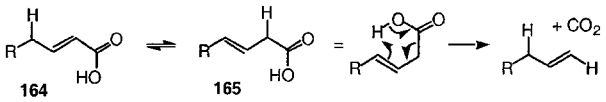


E. Proton Exchange

Amides, RCONHR′, will exchange their NH hydrogens with water. The kinetics can be followed with isotopes or with NMR techniques that distinguish NH protons from OH. The reaction is catalyzed by both H^+ and OH^-. The mechanism of the base-catalyzed reaction is simply removal of the NH, to create the amide's conjugate base as an intermediate that next abstracts a different H from water. The mechanism of the acid-catalyzed reaction was elucidated with various $RCONH_2$ (**162**). Advantage was taken of the partial double-bond character (**43**, R′ = H), which maintains H in two different sites (labeled H_E and H_Z in **162**), each with its own NMR signal. For electron-donating R the exchange was found to occur simply by protonation

on nitrogen, to produce $RC(=O)NH_3^+$. In this intermediate there is no longer any partial double-bond character to impede C–N rotation. Then when one of the three hydrogens is returned to solvent, one possibility is that the original H_E may remain in the amide but be transferred to the H_Z site, to produce **163**. For electron-donating R the exchange was found to occur by a more circuitous mechanism. Protonation on oxygen forms $RC(-OH)=NH_2^+$, which can lose either of the two NH to produce two stereoisomeric RC(–OH)=NH as intermediates. Reversal of these two steps then regenerates the amide, but with one of its NH groups exchanged. However, those NH exchange only with solvent, without exchanging with each other. Thus the two mechanisms could be distinguished by comparing intramolecular exchange with intermolecular, since only via $RC(=O)NH_3^+$ is there opportunity for intramolecular exchange.



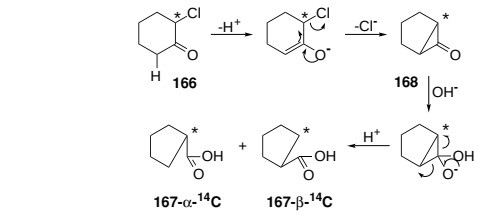


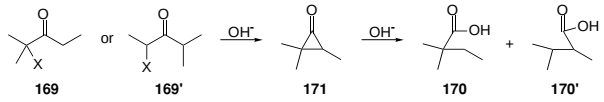
F. Decarboxylation

α,β-Unsaturated carboxylic acids, RCH=CHCOOH, decarboxylate to alkenes, $RCH=CH_2$, plus CO_2. Yet $(CH_3)_3CCH=CHCOOH$ does not show this reactivity. This is a clue that a γ hydrogen is required. The accepted mechanism involves acid-catalyzed isomerization of the α,β-unsaturated acid (**164**) to the β,γ-unsaturated acid (**165**), which can undergo a concerted decarboxylation. Three further pieces of evidence are that α,β-unsaturated acids do isomerize to the β,γ-unsaturated acid faster than they decarboxylate, β,γ-unsaturated acids do undergo decarboxylation with isomerization (similar to the concerted decarboxylation of β-keto acids, $RCOCH_2COOH$), and deuterium is incorporated into the γ position when the reaction is carried out with acid–OD.

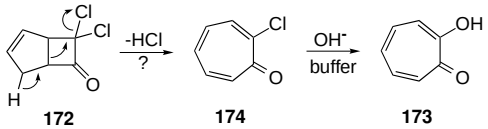
G. α-Haloketones

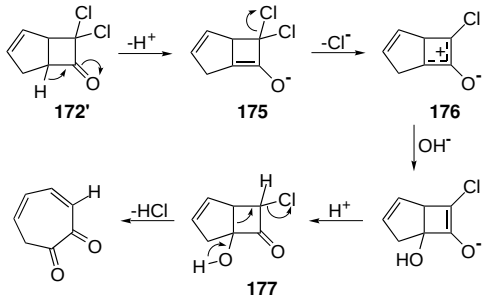
The Favorskii rearrangement is the conversion of an α-halo ketone, RR′CX–COR″, in alkali to a carboxylic acid, RR′R″C–COOH (or ester). Reaction of 2-chlorocyclohexanone-2-^{14}C (**166**, $*=^{14}C$) produces cyclopentanecarboxylic acid (**167**) with the label distributed equally between Cα and Cβ. This is evidence for a symmetric intermediate, the cyclopropanone **168**, which can open in two ways. Similarly, reaction of 2-halo-2-methyl-3-pentanone or 2-halo-4-methyl-3-pentanone (**169** or **169′**, X = Cl or Br) with sodium hydroxide gave the same 4:1 mixture of 2,2- and 2,3-dimethylbutanoic acids (**170**, **170′**) from any of the four different reactants. This is evidence for reaction via trimethylcyclopropanone (**171**) as intermediate. Moreover, when that intermediate was synthesized independently [from dimethylketene, $(CH_3)_2C=C=O$, and diazoethane, $CH_3CH=N^+=N^-$], it produced that same 4:1 mixture when treated with NaOH.





The reaction of **172** in acetate buffer to form **173** shows some similarity to the Favorskii rearrangement. It had been thought to occur by elimination of HCl simultaneous with ring opening to form **174** as a presumed intermediate that is then hydrolyzed further. However, when **174** was prepared independently and subjected to the buffer conditions, it was unreactive. Indeed, the hydrogen that is removed in this mechanism is much less acidic than the adjacent one indicated in **172′**, analogous to **166**. If, instead, that more acidic proton is removed, the resulting enolate anion **175** can lose Cl^-. The resulting **176** is isomeric to a cyclopropanone, like **168** or **171**, but with an open zwitterionic structure to avoid angle strain. Addition of hydroxide followed by protonation produces **177**. This next loses HCl and opens the central ring to form a tautomer of the final **173**. Support for this mechanism was obtained by labeling the CCl_2 carbon and showing that it acquires H.





VI. QUANTITATIVE RELATIONSHIPS

A. Brønsted Relationship

Much insight has come from recognizing quantitative regularities among the phenomena of organic chemistry. This is simply a highly developed method for reasoning by analogy. It is especially powerful since it permits the extension of understanding from one class of reactions to another.

The earliest case was the treatment of reactions subject to general acid or general base catalysis [Eq. (51), adapted from Eq. (39)]. The rate constant k_{HA} reflects the speed with which acid HA can act as a proton donor, and k_B reflects the speed with which base B can act as a proton acceptor. There ought to be a relationship between the kinetic power of an acid or base and its thermodynamic acidity or basicity: k_{HA} ought to increase as HA becomes a stronger acid, and k_B ought to increase with the basicity of B. A quantitative expression of this relationship is Eq. (52), where c is a constant and where the basicity constant of B has been replaced by the acidity constant of its conjugate acid BH^+:

$$k_{obs} = k_H[H^+] + k_{HA}[HA]$$

$$\text{or} \quad k_{obs} = k_{OH}[OH^-] + k_B[B^-], \qquad (51)$$

$$\log_{10} k_{HA} = -\alpha p K_a^{HA} + c$$

$$\text{or} \quad \log_{10} k_B = \beta p K_a^{BH+} + c. \qquad (52)$$

Therefore if k_{HA} or k_B is measured for a series of acids or bases and its logarithms plotted against $-pK_a^{HA}$ or $pK_a^{BH^+}$, a straight line should result with slope α or β. Such an equation is known as the Brønsted relationship.

For example, the acid hydrolysis of ethyl vinyl ether, $C_2H_5OCH{=}CH_2$, proceeds by rate-limiting proton transfer to produce $C_2H_5O^+{=}CH{-}CH_3$, which then reacts rapidly with water and eventually cleaves to C_2H_5OH and $CH_3CH{=}O$. Each acid has its own rate constant k_{HA} for proton transfer; a plot of $\log_{10} k_{HA}$ versus $-pK_a$ of that acid is shown in Fig. 18. The slope α is 0.68 and the intercept c (at $pK_a = 0$) is 0.38.

B. Linear Free Energy Relationships

Figure 18 is an example of a linear free energy relationship. The horizontal axis is related to the free energy of acid dissociation via Eq. (14) and the vertical axis is related to the free energy of activation via Eq. (23). A model for such behavior can be obtained by constructing the reaction coordinate as a composite of the dissociation of the reactant O–H bond and the formation of the product C–H bond, as indicated in Fig. 19. The transition state may be taken as the point where the two curves cross, when bond formation begins to compensate for the bond breaking.

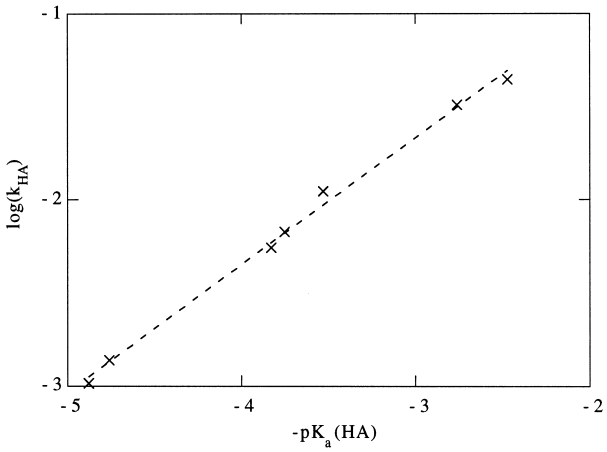


FIGURE 18 Brønsted plot for the hydrolysis of ethyl vinyl ether. Acids RCOOH, from left to right, are $R = NCCH_2$, $ClCH_2$, CH_3OCH_2, H, $HOCH_2$, CH_3, and CH_3CH_2.

The activation energy is indicated as $\Delta G^{\ddagger}$ and the free energy of the reaction as ΔG°. What happens if the product is modified so as to become less stable, as indicated by the dashed curve? The activation energy is now $\Delta G^{\ddagger\prime}$ (distinguished with a prime), higher than $\Delta G^{\ddagger}$, and the free energy of reaction is $\Delta G^{\circ\prime}$, less negative than ΔG°. However, the change of activation energy, $\Delta G^{\ddagger\prime} - \Delta G^{\ddagger}$, is less than $\Delta G^{\circ\prime} - \Delta G^{\circ}$, as can be seen from the vertical distances between the two sets of arrow tips. This result can be written in the following form, where α is between 0 and 1:

$$\Delta G^{\ddagger\prime} - \Delta G^{\ddagger} = \alpha(\Delta G^{\circ\prime} - \Delta G^{\circ}). \qquad (53)$$

It can be shown that Eq. (52) is an example of this equation.

The slope α is a selectivity parameter. If α is small, then all reactions proceed at nearly the same rate, without

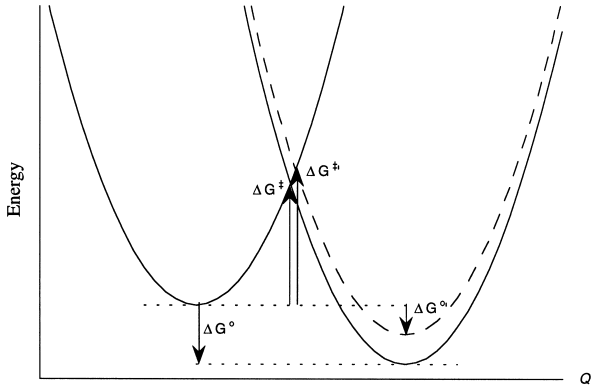
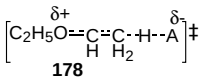


FIGURE 19 Variation of energy along a reaction coordinate constructed from cleaving the reactant bond and forming the product bond. Solid curve is for the standard product, the dashed curve for the product denoted with a prime.

selection for thermodynamic favorability. If α is near 1, then the thermodynamic stabilities fully influence the activation energy. Thus α is a quantitative measure of product-development control, or the extent to which the transition state partakes of the stability of the product. In a very approximate sense, α represents a measure of the extent to which the transition state resembles the product. For example, in the above hydrolysis of ethyl vinyl ether, the α of 0.68 can be identified loosely with the partial negative charge $\delta-$ acquired by A in the transition state (**178**).



178

If instead the product becomes more and more stable, the product curve in Fig. 19 would drop, and its intersection with the reactant curve would also drop in energy. Thus the activation energy is lowered, opposite to the change above, where the product became less stable. Moreover, as the product curve drops, its intersection with the reactant curve moves back (to the left) along the reaction coordinate. Thus the transition state resembles the reactant more and more. There is less and less product-development control, and α diminishes. In the extreme case of a very stable product the reaction becomes fast but unselective.

Some of these conclusions can be seen in the reactions of some hydrocarbons with radical species. The relative rate constants of hydrogen abstraction [Eq. (54)] by chlorine atoms, bromine atoms, trichloromethyl radicals, and under conditions of *N*-bromosuccinimide (NBS) bromination are listed in Table X. Also included are the C—H bond-dissociation energies. For each radical the rates tend to increase as the BDE decreases. This is product-development control, arising because the weaker the bond, the faster it is cleaved. Each of these is also a linear free energy relationship, although the linearity is not perfect because phenyls and methyls have different steric requirements. The chlorination reaction is rather unselective, consistent with a very stable product, owing to the strength of the H—Cl bond that is formed (BDE = 103 kcal/mole, compared to 87 kcal/mole for H—Br). The other reactions are more selective, especially with $Cl_3C\cdot$. Furthermore, the selectivity with NBS is the same, within experimental error, as with Br·. It would be very fortuitous for the *N*-bromosuccinimidyl radical to have the same selectivity as Br·. Therefore it was concluded that the reactive radical in NBS brominations is Br· and not *N*-bromosuccinimidyl.

$$\mathrm{R{-}H} + \mathrm{X}\cdot \rightarrow \mathrm{R}\cdot + \mathrm{H{-}X}. \tag{54}$$

C. Hammett Equation

There is no requirement that the $\Delta G^{\ddagger}$ and ΔG° in Eq. (53) must refer to rates and equilibria of the same reaction. The following Hammett equation is an extension to the comparison of substituent effects in two different reactions:

$$\log_{10} k_{\mathrm{X}} = \rho\sigma_{\mathrm{X}} + c. \tag{55}$$

Here the substituent constant σ_{X} is defined by the following equation from the measured acid-dissociation constants of substituted and unsubstituted benzoic acids, and k_{X} and k_{H} are rate constants for substituted and unsubstituted reactants:

$$\begin{aligned}\sigma_{\mathrm{X}} &= \log_{10}\left(K_{\mathrm{a}}^{\mathrm{XC_6H_4COOH}} / K_{\mathrm{a}}^{\mathrm{C_6H_5COOH}}\right) \\ &= \mathrm{p}K_{\mathrm{a}}^{\mathrm{C_6H_5COOH}} - \mathrm{p}K_{\mathrm{a}}^{\mathrm{XC_6H_4COOH}}.\end{aligned} \tag{56}$$

Table XI is a short list of substituent constants. Increasingly positive values correspond to greater electron-withdrawing power, which stabilizes the negative charge of the carboxylate anion and thus increases the acidity. In practice, rate constants k_{X} for a series of reactants with different substituents X are plotted against σ_{X} to obtain a straight or nearly straight line whose slope is the reaction

TABLE X Rate Constants of C—H Abstraction (per H, Relative to Toluene) by Cl·, Br·, $Cl_3C\cdot$, and with *N*-Bromosuccinimide (NBS)

R	k/k_{tol}(Cl)	k/k_{tol}(Br)	k/k_{tol}(Cl_3C)	k/k_{tol}(NBS)	BDE(R—H)
CH_3CH_2	0.77	1.6×10^{-5}			98
$(CH_3)_2CH$	3.3	3.4×10^{-3}			95
$(CH_3)_3C$	4.6	0.30			92
$PhCH_2$	=1	=1	=1	=1	90
$PhCH(CH_3)$	2.5	17	50	21	87
Ph_2CH	2.0	9.6	50	10	85
$PhC(CH_3)_2$	5.6	37	260	45	85
$Ph_2C(CH_3)$		42	650		84
Ph_3C	7.3	18	160		83

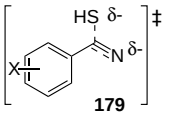
TABLE XI Substituent Constants Derived from pK_a of Benzoic Acids, XC_6H_4COOH

X	σ_X	X	σ_X
p-MeO	−0.28	*p*-Br	0.23
p-Me	−0.17	*m*-Br	0.39
m-Me	−0.06	*p*-NC	0.70
H	= 0.00	*m*-O_2N	0.71
m-MeO	0.11	*p*-O_2N	0.78

constant ρ and whose intercept is c (which equals $\log_{10} k_H$ in an ideal case).

Figure 20 shows such a plot for the reaction of substituted benzonitriles with HS^-:

$$XC_6H_4CN + HS^- \rightarrow XC_6H_4C(=S)NH_2. \qquad (57)$$



179

The slope is 2.1. The positive value reflects the fact that electron-withdrawing substituents accelerate the reaction. This is consistent with a transition state **179** that carries a negative charge. This transition state is not the same as a carboxylate anion, but it is analogous, in that any substituent that stabilizes the carboxylate also stabilizes the transition state. If the slope had been negative, as in some reactions, it would mean that electron-donating substituents accelerate the reaction, consistent with the development of a positive charge in the transition state.

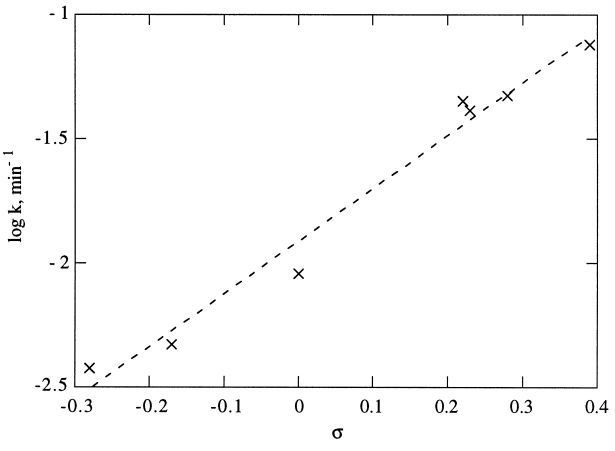


FIGURE 20 Hammett plot for reaction of XC_6H_4CN with HS^-. Substituents X from left to right are *p*-CH_3O, *p*-CH_3, H, *p*-I, *p*-Br, *p*-Cl, and *m*-Br.

For many reactions it is possible to recognize an analogy either to acid dissociation of substituted benzoic acids or to other model reactions. Among these models are acidities of phenols, XC_6H_4OH, or aliphatic acids, XCH_2COOH, rates of solvolysis of $XC_6H_4C(CH_3)_2Cl$, and even the ability of a solvent to stabilize the ground electronic state of a dyestuff. Then the effects of substituents on both equilibrium constants and rate constants can be correlated quantitatively with the substituent effects in those model reactions. No analogy is perfect, but one may provide an understanding of a new reaction in terms of a more familiar one.

SEE ALSO THE FOLLOWING ARTICLES

CHEMICAL KINETICS, EXPERIMENTATION • CHEMICAL THERMODYNAMICS • HYDROGEN BOND • KINETICS (CHEMISTRY) • ORGANIC CHEMICAL SYSTEMS, THEORY • ORGANIC CHEMISTRY, SYNTHESIS • ORGANOMETALLIC CHEMISTRY • PERIODIC TABLE (CHEMISTRY) • PHOTOCHEMISTRY, MOLECULAR • POTENTIAL ENERGY SURFACES • QUANTUM CHEMISTRY • STEREOCHEMISTRY

BIBLIOGRAPHY

BERNASCONI, C. F. (ED.). (1986). "INVESTIGATION OF RATES AND MECHANISMS OF REACTIONS," 4TH ED., WILEY, NEW YORK.

CAREY, F. A., AND SUNDBERG, R. J. (2000/1990). "ADVANCED ORGANIC CHEMISTRY," PART A, 4TH ED. (2000), PART B, 3RD ED. (1990), KLUWER/PLENUM, NEW YORK.

CARROLL, F. A. (1998). "PERSPECTIVES ON STRUCTURE AND MECHANISM IN ORGANIC CHEMISTRY," BROOKS/COLE, PACIFIC GROVE, CA.

LOWRY, T. H., AND RICHARDSON, K. S. (1987). "MECHANISM AND THEORY IN ORGANIC CHEMISTRY," 3RD ED., HARPER & ROW, NEW YORK.

MARCH, J. (1992). "ADVANCED ORGANIC CHEMISTRY: REACTIONS, MECHANISMS, AND STRUCTURE," 4TH ED., WILEY, NEW YORK.

MASKILL, H. (1985). "THE PHYSICAL BASIS OF ORGANIC CHEMISTRY," OXFORD UNIVERSITY PRESS, NEW YORK.

MILLER, B. (1998). "ADVANCED ORGANIC CHEMISTRY: REACTIONS AND MECHANISMS," PRENTICE HALL, UPPER SADDLE RIVER, NJ.

MULLER, P. (ED.). (1994). "GLOSSARY OF TERMS USED IN PHYSICAL ORGANIC CHEMISTRY," *Pure Appl. Chem.* **66,** 1077–1184 (HTTP://WWW.CHEM.QMW.AC.UK/IUPAC/GTPOC/).

PROSS, A. (1995). "THEORETICAL AND PHYSICAL PRINCIPLES OF ORGANIC REACTIVITY," WILEY, NEW YORK.

STOWELL, J. C. (1994). "INTERMEDIATE ORGANIC CHEMISTRY," WILEY, NEW YORK.

TIDWELL, T. T., RAPPOPORT, Z., AND PERRIN, C. L. (EDS.). (1997). "PHYSICAL ORGANIC CHEMISTRY FOR THE 21ST CENTURY (A SYMPOSIUM IN PRINT)." *Pure Appl. Chem.* **69,** 210–292 (HTTP://WWW.IUPAC.ORG/PUBLICATIONS/PAC/SPECIAL/0297/INDEX.HTML).

WILLIAMS, A. (2000). "CONCERTED ORGANIC AND BIO-ORGANIC MECHANISMS," CRC PRESS, BOCA RATON, FL.

Stereochemistry

Ernest L. Eliel
University of North Carolina

GLOSSARY

Angle strain Excess potential energy of a molecule caused by deformation of an angle from the normal, e.g., of a C–C–C angle from the tetrahedral. Also called "Baeyer strain."

Anti or antiperiplanar Conformation of a segment X–C–C–Y in which the torsion angle (which see) is 180° or near 180°. X and Y are said to be anti or antiperiplanar (to each other).

Cahn–Ingold–Prelog (C-I-P) Descriptors Descriptors *R* and *S* (and others) used to describe the spatial arrangement (configuration) of ligands at a chiral center or other chiral element (chiral axis, chiral plane).

Chiral center An atom to which a C-I-P descriptor can be attached, usually a group IV (tetrahedral) or group V (pyramidal) atom. Reflection of the molecule must reverse the descriptor.

Chirality Handedness; the property of molecules or macroscopic objects (such as crystals) of not being superposable with their mirror images. Substances may be called chiral if the constituent molecules are chiral, even if the substance is racemic.

Configuration The spatial arrangement of atoms by which stereoisomers are distinct, but disregarding facile rotation about single bonds.

Conformation The (differing) spatial array of atoms in molecules of given constitution and configuration produced by facile rotation about single bonds.

Conformers Stable conformations (located at energy minima). Isomers which differ by virtue of facile rotation about single bonds, usually readily interconvertible.

Constitution (of a molecule) The nature of its constituent atoms and their connectivity.

Diastereomers (diastereoisomers) Stereoisomers that are not mirror images of each other.

Eclipsed conformation Conformation of a nonlinear array of four atoms X–C–C–Y in which the torsion angle (which see) is zero or near zero (also "synperiplanar"). X and Y are said to be eclipsed.

Enantiomer(ic) excess In a partially or completely

resolved substance, the excess of one enantiomer over the other as a percentage of the total.

Enantiomers Stereoisomers that are mirror images of each other.

Enantiomorphs Macroscopic objects (e.g., crystals) whose mirror images are not superposable with the original entity.

Factorization The analysis, where possible, of chirality in terms of individual chiral elements, such as chiral centers, chiral axes, and chiral planes.

Gauche Conformation of a segment X–C–C–Y in which the torsion angle (which see) is $\pm 60^\circ$ or near $\pm 60^\circ$ (also "synclinal"). X and Y are said to be gauche (to each other).

Kinetic resolution Separation of enantiomers by virtue of their unequal rates of reaction with a nonracemic chiral reagent.

Optical activity The property of chiral assemblies of molecules or chiral crystals of rotating the plane of polarized light.

Optical purity The ratio of the observed specific rotation of a substance to the specific rotation of the enantiomerically and chemically pure substance.

Optical rotation Rotation of the plane of polarized light produced by the presence of chiral molecules or chiral crystals in the light path.

Racemic mixture (racemate) A composite of equal amounts of opposite enantiomers.

Racemization Conversion of individual enantiomers into a racemic mixture.

Resolution Separation of enantiomers from a racemic mixture.

Staggered conformation Conformation of a nonlinear array of four atoms (A–B–C–D) in which the torsion angle (which see) is near 60°. Ligands A and D are said to be *gauche* or *synclinal* to each other.

Stereochemistry Chemistry in three dimensions; topographical aspects of chemistry.

Stereogenic center An atom where exchange of two of its ligands changes the configuration. Such an atom is usually, but not necessarily, a chiral center.

Stereoisomers Isomers of the same constitution but differing in spatial arrangement of their constituting atoms.

Torsion angle In a nonlinear array of four atoms A–B–C–D, the angle between the plane containing A, B, and C and the plane containing B, C, and D.

Torsional strain Excess potential energy of a molecule caused by incomplete staggering of pertinent bonds (e.g., deviation of the torsion angle in H–C–C–H from the normal 60°). Also called "Pitzer strain."

van der Waals or compression strain Excess potential energy of a molecule caused by approach of two or more of its constituting atoms within the repulsive regime of the van der Waals potential. Also called "nonbonded (repulsive) interaction."

I. INTRODUCTION

This article deals with the stereochemistry of organic compounds, although many of the general principles also apply to organometallic and inorganic compounds.

The term "stereochemistry" is derived from the Greek "stereos" meaning solid—it refers to chemistry in three dimensions. Since nearly all organic molecules are three dimensional (with the exception of some olefins and aromatics to be discussed later), stereochemistry cannot be considered a branch of chemistry. Rather it is an aspect of all chemistry, or, to put it differently, a point of view which has become increasingly important and, as will be shown, essential for the understanding of chemical structure and function.

II. HISTORY. OPTICAL ACTIVITY

Historically, stereochemical thinking developed from the observations of J. B. Biot in 1812–1815 that quartz crystals as well as solutions of certain organic substances rotate the plane of polarized light; this phenomenon is called "optical rotation." When the beam of light coming toward the observer rotates the plane to the right (clockwise) we call the rotation positive (+); when the rotation is to the left or counterclockwise, it is negative (−). The angle of rotation α is proportional to the concentration c (conventionally expressed in grams per milliliter, i.e., density, for solids and pure liquids and gases) and the thickness of the layer—usually the length of the cell in which the liquid or solution is contained (conventionally expressed in decimeters): $\alpha = [\alpha] \cdot l \cdot c$ ("Biot's law"). The proportionality constant $[\alpha]$ is called "specific rotation": $[\alpha] = \alpha / l\,(\mathrm{dm}) \cdot c(\mathrm{g/cm^3})$. For solutions, where the concentration c' is conventionally expressed in g/100 ml, $[\alpha] = 100\alpha / l\,(\mathrm{dm}) \cdot c'\,(\mathrm{g/100\ ml})$. The constant $[\alpha]$ (conventionally reported without units) is used for the characterization of "optically active" substances (i.e., which display optical rotation). It varies with temperature and with the wavelength of the light used in the observation, which need to be indicated (as subscripts and superscripts, respectively): thus, $[\alpha]^t_\lambda$, where t is the temperature in $^\circ$C and λ is the wavelength in nanometers (nm). Unfortunately, since most substances are prone to solvation and intermolecular association in solution or in the liquid state, and since such association varies with concentration and solvent, these items must also be recorded. A

typical specific rotation might thus read $[\alpha]_D^{20} + 57.3 \pm 0.2$ (95% EtOH, $c = 2.3$), denoting measurement at 20°C at the sodium D line (589 nm) in 95% ethanol at a concentration of 2.3 g/100 ml. (Monochromatic light of this wavelength is easily generated in a sodium vapor lamp and has thus classically been used for polarimetric measurements.)

In 1822 Sir John Herschel found that mirror-image crystals of quartz (discovered by R. J. Haüy in 1801 and called "enantiomorphs") rotate the plane of polarization in opposite directions. This provided the first correlation of optical rotation with enantiomorphism, in this case of crystals. In 1848 Louis Pasteur achieved separation of the enantiomorphous crystals he detected in the sodium–ammonium salt of the (optically inactive) paratartaric acid (today called racemic tartaric acid, see below) and thereby obtained two different substances, one of which rotated polarized light to the right, the other to the left, even in aqueous solution. Pasteur concluded that the enantiomorphous crystals were made up of molecules that themselves differed as object and (reflection) image on the molecular scale. Such mirror-image molecules are called "enantiomers" and the separation which Pasteur accomplished is called "resolution." Since the difference between enantiomers resembles the difference between a right and a left hand (which are also mirror images of each other, but otherwise essentially identical in their dimensions), we call such molecules "chiral" (Greek "cheir" = hand) and we call the property of certain molecules to display enantiomerism "chirality" (terms coined by Lord Kelvin in 1893). While every molecule has a mirror image, chirality exists only if the image is nonsuperposable with the original molecule, just as a right shoe is not superposable with a left one. (In contrast, socks worn on right and left feet are superposable; they are "achiral.")

III. STEREOISOMERISM

The understanding of molecular structure was not well enough advanced in 1848 for Pasteur to explain enantiomerism (chirality) in terms of atomic arrangement. The basis for that was laid only a decade or so later when, in separate publications, A. S. Couper, F. A. Kekulé, J. Loschmidt, and A. Crum Brown illuminated the structure of molecules in terms of the connectivity between their constituent atoms. Then, in 1874, J. H. van't Hoff in the Netherlands and J. A. Le Bel in France simultaneously proposed the structural basis for chirality: When four different atoms or groups (jointly called "ligands") are attached *tetrahedrally* to a given carbon atom, two mirror-image arrangements are possible (Fig. 1) corresponding to the two enantiomers. The enantiomers are said

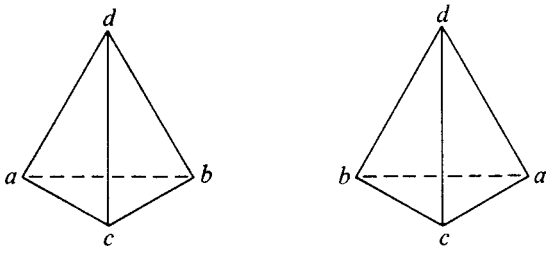


FIGURE 1 Tetrahedral molecule Cabcd (the carbon atom at the center of the tetrahedron is conventionally not shown). [Reprinted with permission from Eliel, E. L., and Wilen, S. H. (1994). "Stereochemistry of Organic Compounds," Wiley, New York.]

to differ in "configuration." Thus they explained not only the stereoisomerism of such simple molecules as CHFBrI or $CH_3CHOHCO_2H$, but also that of more complex molecules such as tartaric acid (Fig. 2). The two tartaric acids, A and B in Fig. 2 (separated by Pasteur as sodium–ammonium salts), are enantiomers. Pasteur's starting salt came from a 1:1 mixture of the two, called "racemic tartaric acid." A "racemate" is a mixture of equal quantities of corresponding enantiomers. Since the numerical values for the optical rotations of the two enantiomers (negative, or −, for one; positive, or +, for the other) are equal and opposite, the racemate displays no optical activity.

In later work Pasteur discovered a fourth species of tartaric acid [if we count the enantiomers A and B and the racemate (A + B) as three distinct species]. He was not able to explain the nature of this optically inactive isomer, but the Le Bel–van't Hoff theory led to assignment of the structure C (Fig. 2) to this stereoisomer. It has the same connectivity (constitution) as A and B but differs in configuration. The reason for its lack of optical activity is that it is superposable with its mirror image D; i.e., it is achiral. It is a stereoisomer of A and B, but not an enantiomer. Such stereoisomers that are not mirror images of each other are called "diastereomers"; thus C is a diastereomer of A and of B (and vice versa), whereas A is an enantiomer of B (and vice versa). Compound C is called "meso-tartaric acid," the prefix "meso" indicating that it is the achiral member in a set of diastereomers that also contains chiral members.

IV. STEREOISOMERISM OF ALKENES AND CYCLANES

So far it would appear that stereoisomerism is dependent on three-dimensional structure, but van't Hoff recognized that stereoisomers can also exist in two dimensions, as in *cis*- and *trans*-disubstituted ethylenes (Fig. 3). The substituents may be on the same side (*cis*) or on opposite sides (*trans*) at the two ends of the (planar) double bond.

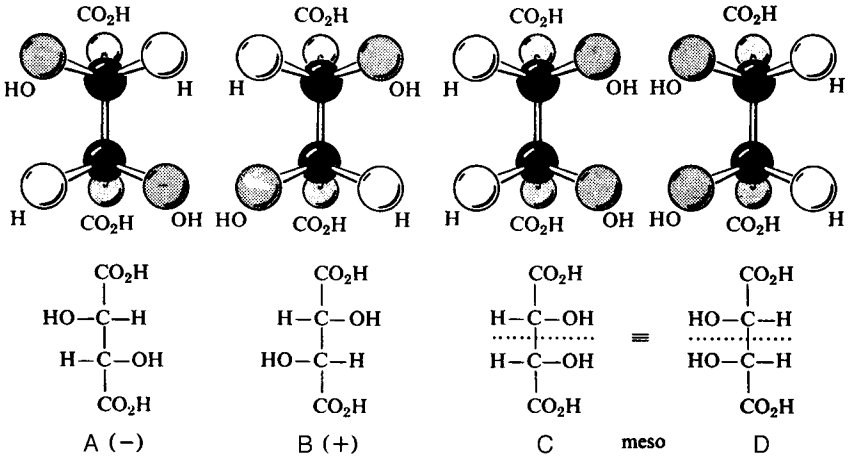


FIGURE 2 The tartaric acids. [Reprinted with permission from Eliel, E. L., and Wilen, S. H. (1994). "Stereochemistry of Organic Compounds," Wiley, New York.]

The classical *cis/trans* nomenclature for alkenes works well for 1,2-disubstitued ethenes (A, B), but with tri- (C) and tetrasubstituted species it is not unequivocal. Thus in the propenoic acid (C), Br is *trans* to CO_2H but *cis* to Cl. For an unequivocal description, substituents at the same terminus are ordered by the Cahn–Ingold–Prelog system (see below; for the present purpose it suffices to recognize that substituents of higher atomic number have priority over those of lower atomic number). Descriptors in this currently used system, are *Z* (for the German *zusammen*) if the higher priority ligands are on the same side of the double-bond system and *E* (for *entgegen*) when they are on opposite sides. Thus C in Fig. 3 is (*Z*)-2-chloro-3-bromopropenoic acid.

Cis–trans isomerism is also found in cyclanes, which, for the purpose of counting stereoisomers, may be considered planar (but see below). Figure 4 shows the *cis–trans* isomerism of 1,2- (A–C) and 1,3- (D, E) dichlorocyclobutanes. The situation in the 1,2 isomers is similar to that in the tartaric acids: there are two enantiomers (A, B; *trans*) plus an (achiral) meso isomer (C; *cis*), which is a diastereomer of A and B. The situation in the 1,3 isomers (D, *cis*; E, *trans*) is different. While D and E are diastereomers, neither of them is chiral (each one is superposable with its own mirror image). Carbons 1 and 3 are not chiral centers since there is no chirality in the molecule, and yet, changing their relative position (*cis* or *trans*) gives rise to (dia)stereoisomers. Carbons 1 and 3 are therefore called "stereogenic." All chiral centers are stereogenic, but, as seen in this case, not all stereogenic centers are chiral centers. The *E*/*Z* system is *not* used for cycloalkanes.

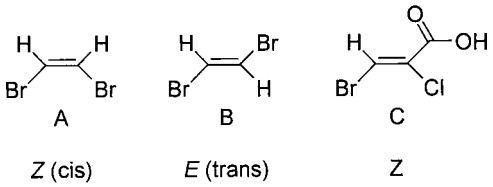


FIGURE 3 The *E*, *Z* nomenclature for *cis* and *trans* substituted alkenes.

V. PROPERTIES OF STEREOISOMERS

Enantiomers, though not superposable, resemble each other very closely (as do right and left hands). The distances (both bonded and nonbonded) between corresponding constituent atoms are the same, and thus

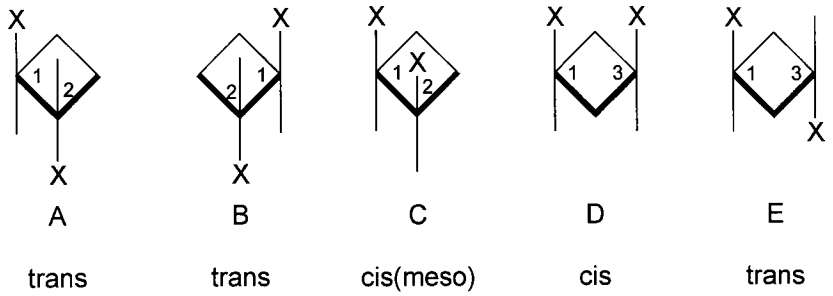


FIGURE 4 Stereoisomers of 1,2- and 1,3-disubstituted cyclobutanes.

enantiomeric species have been called "isometric." In the absence of a second source of chirality their physical and chemical properties are identical. This is true of melting and boiling points, density, refractive index, a whole host of spectral properties (UV, IR, NMR, mass spectrum, etc.), thermodynamic properties, chromatographic behavior, and reactivity with achiral reagents (thus, enantiomeric esters are hydrolyzed in the presence of achiral acids or bases at the same rate.) A macroscopic analogy is that of right and left feet (chiral) which fit equally into either one of a pair of socks (achiral).

However, when the chemical or physical agents are themselves chiral, this equivalence breaks down, just as left and right feet do not fit equally well into a right shoe. Examples are esterification of racemic α-methyl benzyl alcohol, $C_6H_5CHOHCH_3$, with the 2,4-dichlorophenyl ether of (+)-lactic acid, $Cl_2C_6H_3OCH(CH_3)CO_2H$ [the (−) enantiomer reacts faster], irradiation of racemic α-azidopropanoic acid dimethylamide, $CH_3CH(N_3)CON(CH_3)_2$, with left-circularly polarized light (chiral; see below), which leads to faster photochemical destruction of the (+) enantiomer and chromatography on chiral stationary phases.

The situation is otherwise with diastereomers, which are not isometric and therefore do not have identical physical or chemical properties even in the absence of external chirality. Thus, meso-tartaric acid (C in Fig. 2) melts at 140°C, whereas the diastereomeric (+) and (−) acids A and B both melt at 170°C (as explained above, A and B have the same melting point even though their optical rotations are opposite).

VI. SIGNIFICANCE

Optical rotation is useful to characterize enantiomers of chiral substances and has played an important historical role in the development of stereochemistry. However, the real significance of stereoisomerism lies in the concept of fit or misfit, epitomized by the fit of a right glove with a right hand and the corresponding misfit of a left glove. In 1858, Pasteur discovered that, when a solution of racemic tartaric acid is inoculated with the mold *Penicillium glaucum*, the (+) acid is metabolized (and thereby destroyed) by the microorganism and the (−) acid remains behind. This selectivity is due to the organism's enzymes being able to interact with the naturally occurring (+) enantiomer, but not with the other (−). A further example is the hydrolysis of the acetyl derivative of a racemic α-amino acid, $RCH(NHCOCH_3)CO_2H$, promoted by the enzyme hog kidney acylase. Only the acyl derivative of the (naturally occurring) L-amino acid (see below for notation) is hydrolyzed to L-$RCH(NH_2)CO_2H$, which can then be readily separated from the unhydrolyzed D-$RCH(NHCOCH_3)CO_2H$. The latter can separately be hydrolyzed (with aqueous acid) to D-$RCH(NH_2)CO_2H$; both pure D and L acids can thus be prepared from the racemate. This process is called "kinetic resolution" ("kinetic" because it depends on reaction rate; one enantiomer reacts relatively rapidly, the other much more slowly or not at all). The hand-and-glove analogy applies here: One enantiomer fits into the active site of the acylase enzyme and its hydrolysis is thus promoted; the other enantiomer does not fit, just as a right glove (substrate) fits on a right hand (enzyme analog), whereas a left glove does not fit. Considering the matter more generally, the right-hand/right-glove combination is diastereomeric to the right-hand/left-glove combination; diastereomers have different free energies and so, presumably, have the transition states leading to their formation. The effectiveness of the kinetic separation (resolution) depends on the difference in activation energies (and the associated difference in reaction rate): The greater the difference, the better the resolution. Enzymes tend to excel here because their relatively complex topography usually leads to a large difference in activation energy between enantiomeric substrates (fit vs. misfit in the transition state) and hence a high degree of discrimination.

A similar discrimination is seen in pharmaceuticals, whose action depends on their interaction with enzymes (if they are enzyme inhibitors) or other receptors. By way of example, only the (−) enantiomer of α-methyldopa (α-methyl-3,4-dihydroxyphenylalanine) is an antihypertensive since it is enzymatically converted in the body to the pharmacologically active α-methylnorepinephrine. The (+) enantiomer is inactive. We call the active enantiomer "eutomer" and the inactive one "distomer." Since most drugs have some side effects and therefore should be used at the minimal effective dose, the pharmaceutical industry is much interested in specifically preparing the eutomer, free of the ineffectual and potentially detrimental distomer (which constitutes 50% of the racemate). Other areas where enantiomers differ in function are in the agriculture and flavor industries. Thus in the case of the agrochemical paclobutrazol, the (+) enantiomer is a fungicide, whereas the (−) enantiomer is a plant growth regulator. Concerning flavor, the (−) enantiomer of carvone (2-methyl-5-isopropylidenecyclohexanone) has the odor of spearmint, whereas its (+) enantiomer has the odor of caraway.

VII. SEPARATION (RESOLUTION) AND RACEMIZATION OF ENANTIOMERS

Since enantiomers are so similar to each other, it is not surprising that their separation requires special methodology, which can be summarized as follows:

1. Separation by crystallization (Pasteur's first method, 1848)
2. Separation by formation and separation of diastereomers (Pasteur's second method, 1853): (a) separation by crystallization, (b) separation by chromatography
3. Asymmetric transformation (a) of diastereomers, (b) of enantiomers
4. Kinetic resolution: (a) chemical, (b) enzymatic (Pasteur, 1858)
5. Separation by chromatography on chiral stationary phases
6. Enantioselective synthesis
7. Synthesis from enantiomerically pure precursors (sometimes called enantiospecific synthesis)
8. Miscellaneous methods

Pasteur's method of manually separating enantiomorphous crystals of enantiomers is obviously not practical; it is also not general. Racemic mixtures can lead to three different kinds of crystals: conglomerates, racemic compounds, and racemic solid solutions. Compounds (where the unit cell, the smallest unit of a crystal, contains equal numbers of enantiomeric molecules) are unsuitable for Pasteurian resolution, as are racemic solid solutions. Only when the enantiomers crystallize in discrete crystals (i.e., as a macroscopic mixture called a "conglomerate") can the two types of crystals be separated even in principle. But only ca. 10% of racemic mixtures (ca. 20% in the case of salts) crystallize as conglomerates. When they do, separation can be achieved by a modification of Pasteur's technique (called the "method of entrainment") involving alternate seeding with one enantiomer, separating the additional crystalline material formed, replenishing the solution with racemate, then seeding with the opposite enantiomer, thereby inducing crystallization of the second enantiomer. Large quantities of enantiomers, for example, of glutamic acid, have been separated in this manner commercially.

Pasteur's second method, formation of diastereomers by reaction of a racemate with an optically active auxiliary or "adjuvant" ("resolving agent"), is much more common. Common resolving agents are naturally occurring chiral acids [such as (−)-malic acid, $HO_2CCHOHCH_2CO_2H$] for chiral bases, and naturally occurring chiral bases, such as (−)-quinine, for chiral acids. The salts formed are often crystalline and can be separated by fractional crystallization. After separation, the resolved acid or base is liberated by treatment of the salt with mineral acid or base, respectively. Alternatively, covalent diastereomers may be formed [e.g., by esterification of a racemic acid with (−)-menthol (2-isopropyl-5-methylcyclohexanol)] and separated by some type of chromatography on an achiral stationary phase. After separation the chiral auxiliary is removed (e.g., by hydrolysis in the case of an ester).

To understand the third method, "asymmetric transformation," we must first take up "racemization," i.e., the conversion of one of the enantiomers into a racemate. This apparently counterproductive process is actually useful: In resolution the undesired enantiomer is produced as an equimolar by-product. Racemization of that enantiomer allows one to start the resolution process over. Actually, racemization involves converting one enantiomer into the opposite one, but since enantiomers have the same free energy, the equilibrium constant between them is unity, i.e., the product of equilibration is the racemate. However, racemization, to be feasible, requires a chemical pathway. For example, a chiral ketone, $RR'CHCOR''$ may be racemized by base via the resonance-stabilized achiral enolate anion $RR'C^-COR'' \Leftrightarrow RR'C{=}CR''O^-$.

If such equilibration occurs concomitant with resolution, it is sometimes possible to convert the racemate entirely into one of the enantiomers. This process, called "crystallization-induced asymmetric transformation," may be observed during crystallization of diastereomers when the stereoisomers to be resolved can simultaneously be equilibrated. An example is phenylglycine, $C_6H_5CH(NH_2)CO_2H$, required as the (−) isomer in manufacture of the antibiotic ampicillin. Equilibration of the enantiomers is effected by adding benzaldehyde to the racemic material (resulting in reversible formation of a Schiff base which is readily racemized) and precipitation of the desired (−) acid as its salt with (+)-tartaric acid. The (+) isomer is concomitantly reconverted to the racemate; in the end, nearly the entire amino acid crystallizes as the (+)-tartrate of the (−) acid.

Kinetic resolution has already been discussed. Purely chemical approaches (exemplified by the resolution of chiral allylic alcohols, e.g., $C_6H_{11}CHOHCH{=}CHCH_3$, with Sharpless' reagent, which contains isopropyl tartrate as the chiral constituent) are currently rare; enzymatic methods (e.g., hydrolysis of esters of chiral alcohols catalyzed by lipase enzymes) are more common since enzymes are frequently highly selective for one enantiomer over the other and the effectiveness of kinetic resolution depends on the degree of selectivity.

While ordinary chromatography does not separate enantiomers (though it can lead to separation of diastereomers), enantiomers can be separated by chromatography employing a chiral stationary phase enriched in a single enantiomer. In that case, the interactions between the two enantiomers of the analyte and the chiral stationary phase are diastereomeric in nature and therefore often differ in strength, the stronger interaction leading to longer retention time.

Asymmetric (or enantioselective) synthesis is a targeted method for obtaining individual enantiomers from achiral precursors. Ordinarily, the introduction of a chiral center (or other chiral element) in the course of synthesis leads to equal formation of the two enantiomers, i.e., to a racemate. Selectivity can, however, be achieved by using a chiral (enantiomeric) reagent or catalyst, in which case the transition states leading to the two enantiomeric products are diastereomeric, or by attaching a chiral auxiliary (see above) to the starting material so that the products are diastereomers rather than enantiomers, in which case their free energies and the activation energies for their formation also differ and one isomer is formed in preference over the other. At the end of the reaction, the chiral auxiliary is chemically removed. By way of example: Reduction of pyruvic acid, CH_3COCO_2H, to lactic acid with an achiral reagent such as sodium borohydride gives racemic lactic acid, $CH_3CHOHCO_2H$, since approach of hydride from either face of the keto-carbonyl group is equally likely. But when reduction (by a reducing coenzyme) is carried out in the presence of the chiral enzyme lactic acid dehydrogenase, only (+)-lactic acid is formed.

A distinct process, sometimes called "enantiospecific synthesis," is one in which an enantiomerically pure starting material (natural or man-made) is converted by standard reactions into an enantiomerically pure product.

Among other methods are diffusion through chiral membranes and partition methods involving chiral solvents.

VIII. STRUCTURE, CONFIGURATION, NOTATION

By "structure" of a molecule we understand the totality of the nature and array of its atoms. This comprises the identity and connectivity of these atoms ("constitution") and their arrangement in space ("configuration," "conformation"; see below). Structure may be determined by X-ray, electron, or neutron diffraction of a crystal of the substance in question.

Constitution can usually be inferred from elemental analysis and chemical degradation or by various spectroscopic methods, such as nuclear magnetic resonance. Constitutional isomers, such as butane and 2-methylpropane, have the same elemental composition but differ in connectivity. In contrast, the two enantiomers of lactic acid, $CH_3CHOHCO_2H$, identical in constitution, differ in the spatial disposition of the ligands at C(2): They are said to differ in "configuration" and may be called configurational isomers. While configuration is a property of the molecule as a whole, it is convenient to "factorize" it into elements of chirality, notably the "center of chirality" [e.g., at C(2) in lactic acid]. This "factorization" allows us to specify configuration by assigning a configurational "descriptor" to each chiral center or other chiral element.

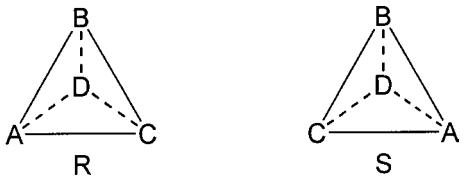


FIGURE 5 Ordering substituents (A > B > C > D) according to the Cahn–Ingold–Prelog chirality rule.

The configurational descriptors now universally used are "*R*" (for *rectus*, right, in Latin) and "*S*" (for *sinister*, left) (Cahn *et al*., 1966); they are sometimes called "CIP descriptors" after their proponents. To assign a descriptor to a chiral center, assume that its four ligands, a, b, c, and d, can be ordered by some convention: a > b > c > d. The molecule is then viewed from the side away from the lowest ranked ligand (d) such that the other three ligands (a, b, c) lie in a plane (Fig. 5). If a–b–c then describe a clockwise array, the descriptor is *R*; if the array is counterclockwise, the descriptor is *S*. To establish priorities in a real molecule one orders the ligands by atomic number, thus in CHFClBr, Br > Cl > F > H. For lactic acid the priority is O > C > H, but no immediate decision is reached for CH_3 versus CO_2H. Here one goes out to the next atom away from the chiral center: O for CO_2H and H for CH_3 and since O > H, CO_2H has priority over CH_3. Where still no decision is reached, one goes out one tier more, thus $–CH_2CH_2OH$ has priority over $–CH_2CH_2CH_2OH$. Once a decision is reached, the process stops. In the outward path, one always gives preference to the atom of higher priority; thus –NHCl has priority over $–N(CH_3)_2$: Cl > C overrides C > H. All ligands on the atom reached must be probed, thus $–CO_2H$ > –CHO (two O's over one O). When the lack of a decision is caused by a doubly bonded ligand (e.g., CH═O vs. CH_2OH as in glyceraldehyde, $HOCH_2CHOHCH{=}O$) the double bond is replaced by two single bonds with the ligands "complemented" at either end; thus –CH═O is considered as O–CH–(OC) and thus has priority over CH_2OH. An absent ligand (as the lone pair in N: or :O:) is considered to have atomic number zero. When chirality is due merely to the presence of an isotope, as in C_6H_5CHDOH, the ligand of higher atomic weight is given priority: O > C > D > H. For more complicated cases, the reader is referred to standard texts.

There are two exceptions to the current use of CIP descriptors, α-amino acids and sugars, where an older nomenclature is often used. Before considering this point, a discussion of projection formulas is required. Since molecules are three dimensional but paper is planar,

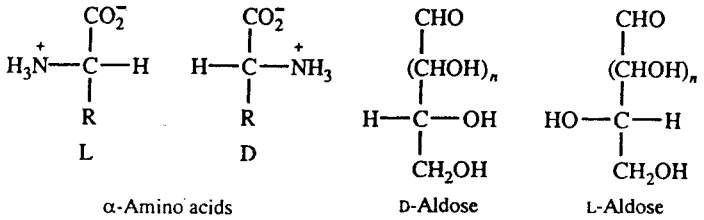


FIGURE 6 The D and L nomenclature for α-amino acids and aldohexoses. [Reprinted with permission from Eliel, E. L., and Wilen, S. H. (1994). "Stereochemistry of Organic Compounds," Wiley, New York.]

several conventions have been developed to represent molecules (or rather their three-dimensional models) in two dimensions. In one of them, the molecule is written with as many atoms as possible in the plane of the paper and with additional atoms being connected with a heavy line (—) when the attached ligand is in front of the plane and with a dashed line (- - -) when it is behind. In another, the so-called "Fischer projection" (after its originator Emil Fischer) the tetrahedron is oriented so that two of the groups (top and bottom) are pointed away from the observer, with the other two (sideways groups) pointing toward the observer, and the molecule is projected in this fashion. Figure 6 shows Fischer projections of α-amino acids and of monosaccharides (aldoses) with the further proviso that the most oxidized group is to be placed at the top and the NH_2 or OH and H ligands on the side. In α-amino acids the symbol D is used when the NH_2 group is on the right, L if it is on the left; it turns out that all naturally occurring α-amino acids are L. In the case of monosaccharides, the chiral center furthest away from the aldehyde or ketone function determines the descriptor and the symbol used is D when the OH is on the right in the Fischer projection formula, L if it is on the left, independent of the configuration of any of the other chiral centers.

The Fischer projection formulas shown above, while useful for the assignment of the descriptors, do not correspond to the actual shape of most molecules. As explained below under the topic of conformation, most molecules exist in "staggered" (Fig. 7) rather than the "eclipsed" conformations implied in Fischer projections. A more realistic representation of, say, (*R*,*R*)-tartaric acid is shown in Fig. 7, which, in addition to the unrealistic three-dimensional formula corresponding to the Fischer projection, displays the staggered conformation in a three-dimensional, so-called "saw-horse" formula and its projection (seen from one end of the molecule) in a so-called Newman projection. [This staggered representation is generated from the eclipsed one by rotation about the C(2)–C(3) bond.]

IX. DETERMINATION OF CONFIGURATION

Since configuration is an integral part of structure, the determination of the architecture of any molecule, naturally occurring or synthetic, is not complete until its configuration is known. For example, the fit of a drug with its receptor or of an inhibitor with an enzyme cannot be understood (or modeled, or rationally improved) absent information on its configuration.

Configuration may be relative or absolute. To say that a right hand fits a right glove is to make a statement of the *relative* configuration of the two. Even a small child may be able to make this correlation. But to recognize that a picture of a glove is that of a right glove in an *absolute* sense is more difficult. The same applies to the determination of the absolute configuration (or sense of chirality) of a molecule.

There are many ways of determining relative configuration (Eliel and Wilen, 1994). The most straightforward one, where accessible, is an X-ray crystal structure; since an X-ray (or neutron or electron) diffraction picture leads to the positions of the constituting atoms in space, their relative orientation (i.e., configuration) can be determined. Relative configuration thus correlates one chiral center

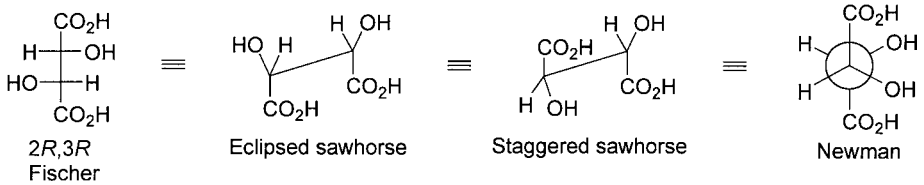


FIGURE 7 Fischer, sawhorse, and Newman representations of (2*R*, 3*R*)-tartaric acid.

with another. The two chiral centers need not be in the same molecule; as will be shown later, configurational correlations between two similar molecules can sometimes be based on comparison of their optical rotations, optical rotatory dispersion, or circular dichroism spectra (see below). It is also possible to tie two chiral centers, one of known absolute configuration, the other unknown, together chemically by either ionic or covalent bonds and determine their relative configuration by X-ray diffraction or other means. Since the absolute configuration of one of the chiral centers is known, that of the other can then be deduced. If the compound containing that center can then be separated by an appropriate chemical reaction, its optical rotation can be measured and thus the necessary correlation between optical rotation ($+$ or $-$) and configuration (*S* or *R*) is established. [There is no general relation whatever between $+$ and $-$ (experimental quantities) and *R* and *S* (descriptors).]

Determination of the absolute configuration (*R* or *S*) of an isolated species is more difficult. Enantiomers are indistinguishable in their physical behavior in scalar measurements (i.e., in measurements not involving absolute orientation in space); ordinary X-ray diffraction is of this type. However J. M. Bijvoet, in the Netherlands, found in 1951 that this impediment can be circumvented by employing X-rays of a wavelength close to the absorption edge of one of the constituent atoms in the molecule to be examined. This specific absorption (usually by a relatively heavy atom, such as sulfur or bromine introduced in the species to be examined by chemical transformation if necessary) leads to a phase shift of the wavefront diffracted by this particular atom. This phase shift causes a pair of spots in the normally centrosymmetric diffraction pattern (so-called "Bijvoet pairs") to become unequal in intensity; from the relative intensity of these spots the absolute configuration of the compound under investigation can be inferred. Thanks to the availability of powerful computers, it has also become increasingly feasible to derive absolute configuration from optical rotation or circular dichroism spectra (see below) by theoretical computation. In some cases, absolute configuration can also be established by examining the crystal habit (macroscopic dimensions) of a crystal in the presence of certain impurities (Addadi *et al*., 1986).

Once the absolute configurations of a few chiral molecules are known, those of others can be established by correlation.

X. CHIRALITY IN ABSENCE OF CHIRAL CENTERS

Although Le Bel's and van't Hoff's understanding of chirality rested on the concept of tetrahedral carbon or, more generally, of what is now called a chiral center, chirality is not dependent on the existence of chiral atoms. Any molecule that is not superposable with its mirror image is chiral. An example, shown in Fig. 8E, is twistane. A secondary criterion for chirality is the absence of a plane of symmetry or a point of inversion. However, chiral molecules may contain simple (proper) axes of symmetry; twistane, in fact, has three mutually perpendicular twofold symmetry axes. One class of chiral molecules already foreseen by van't Hoff (though obtained as individual enantiomers only much later) are appropriately substituted allenes, as shown in Fig. 8A. The orbitals are so disposed that the two double bonds are perpendicular to each other, and so two mutually different substituents at the two termini will give rise to chirality (in contrast to the *cis–trans* isomerism of alkenes, Fig. 3).

Related chiral molecules are appropriately substituted spiranes (Fig. 8B) and alkylidenecycloalkanes (Fig. 8C). These molecules are said to possess chiral axes (along the

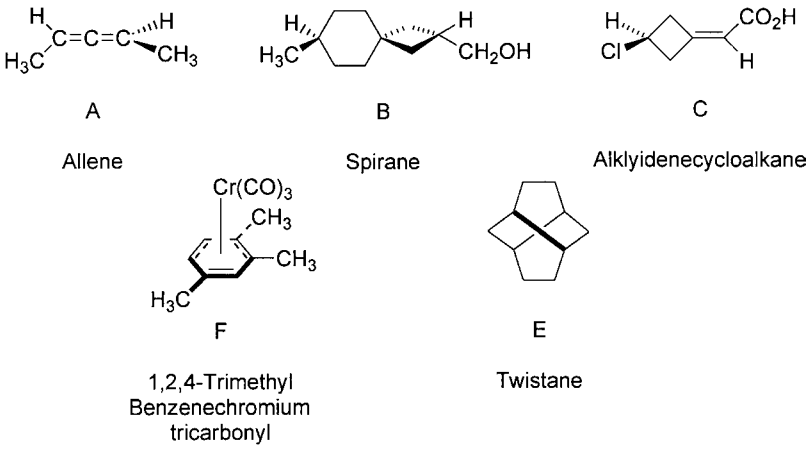


FIGURE 8 Chiral compounds lacking chiral centers.

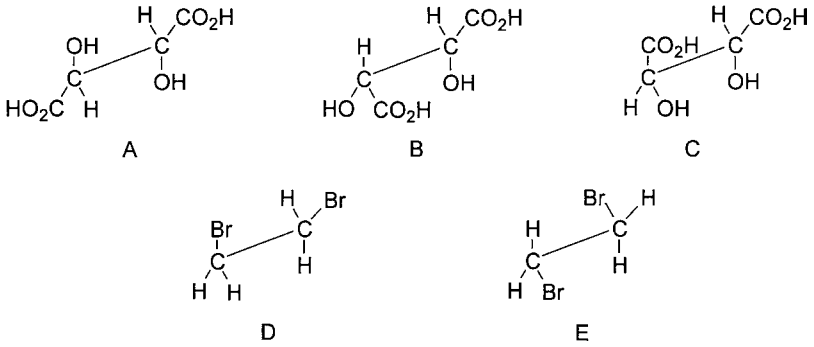


FIGURE 9 Conformers of meso-tartaric acid and of 1,2-dibromoethane.

line of the double bonds or bisecting the rings). Also in this category are the biphenyls, to be discussed later. Yet another common type of chirality due to "chiral planes" is seen in the benzenechromium complex in Fig. 8F. (The normal symmetry plane of the benzene rings is abolished by the out-of-plane coordinated chromium atom.)

XI. CONFORMATION

van't Hoff's counting of stereoisomers was based on the assumption that rotation about single bonds was "free," otherwise there should be at least three isomers of meso-tartaric acid shown in Fig. 9A–C two of which (Fig. 9B, C) would be chiral and enantiomeric. (This is on the assumption that the substituents are staggered—see below—otherwise many more isomers could exist.) In 1932, However, S.-I. Mizushima discovered by vibrational spectroscopy that there are, in fact, two isomers of 1,2-dibromoethane (Fig. 9D, E; vibrational spectroscopy cannot distinguish between enantiomers). It was established later that the bromine substituents are "staggered" rather than "eclipsed," meaning that the torsion angle Br–C–C–Br is 60° (Fig. 9D) or 180° (Fig. 9E) rather than 0°. The structures in Fig. 9D, E are said to differ in "conformation" (rotation about single bonds).

Why are there differing isomers for 1,2-dibromoethane even though they cannot be isolated? The answer to this question was provided by K. S. Pitzer in 1936 when he discovered that ethane itself (Fig. 10) exists in staggered conformation and that the three possible staggered conformations are separated by energy barriers of 12.1 kJ/mole (corresponding to the eclipsed conformation as the energy maximum). Such barriers are high enough to allow detection of the individual conformers by vibrational spectroscopy but far too low to allow chemical separation. What one sees in chemical behavior (and also in physical measurements involving "slow" time scales, such as measurement of dipole moments or of electron diffraction patterns), is an average of the contributing stable conformations (called "conformers") produced by their rapid interconversion.

The staggered conformations of butane (in Newman projections) are shown in Fig. 11 and resemble those in 1,2-dibromoethane (Fig. 9). There are three; in two of them (the enantiomeric *gauche* conformers in Fig. 11A, C the terminal methyl groups are close enough together to give rise to van der Waals repulsion. Thus these conformers are less stable (by 4 kJ/mole) than the anti conformer (Fig. 11B), in which the methyl groups are remote from each other. [The relative instability of the gauche (Fig. 9D) relative to the anti confomer (Fig. 9E) is even greater in $BrCH_2CH_2Br$ because of additional dipole repulsion in Fig. 9D and its enantiomer.] The conformations of Fig. 11A, C in which the torsion angle ω[C(1)–C(2)–C(3)–C(4)] is 60° are called "gauche"

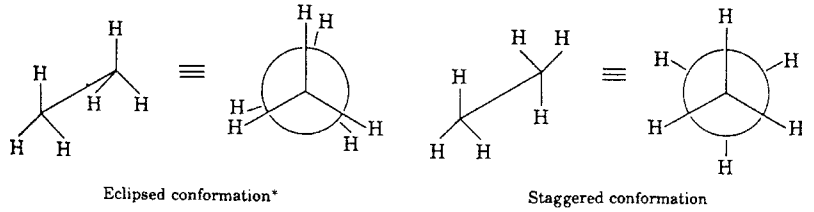


FIGURE 10 Eclipsed and staggered conformations of ethane. Since the hydrogen atoms on the front carbon obscure those in the rear in the eclipsed conformation, the torsion angle is offset by a few degrees in the Newman formula. [Reprinted with permission from Eliel, E., Allinger, N. L., Morrison, G. A., and Angyal, S. J. (1981). "Conformational Analysis," American Chemical Society, Washington, DC. Copyright 1981 American Chemical Society.]

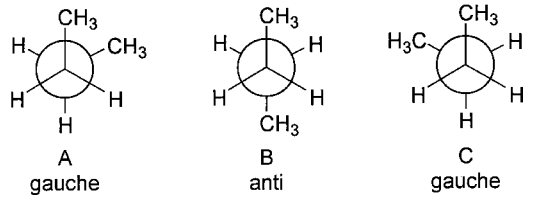


FIGURE 11 Stable conformers of butane.

(French for "skew"). In Fig. 11B the torsion angle is 180°; this conformation is called "anti." Torsion angles often deviate from the ideal values (60° or 180°) for staggered conformations; thus it may be desirable to specify the exact torsion angle ω when known (e.g., from X-ray structure determination). When C(1)–C(2)–C(3)–C(4) describe a right-handed helical turn, ω is positive; for a left-handed turn, it is negative. As an alternative, a system of semiquantitative conformational descriptors more detailed than gauche and anti has been developed by Klyne and Prelog (1960) and is described in the original reference and in Eliel and Wilen (1994).

In straight-chain hydrocarbons larger than butane, rotation about each single bond is possible, giving rise to a large number of conformations; this situation exists especially in linear polymers, where it was studied early by P. Flory. Even when conformations in which the chain coils upon itself in such a way as to generate excessive van der Waals (steric) repulsion are excluded, the number of low-energy conformers will be quite large and a Monte Carlo approach may have to be used to find the family of populated (low-energy) conformers. (Conformers which lie 12 kJ/mole or more above the lowest energy conformer are populated to the extent less than 1% of the total and may be neglected for most purposes.) By way of an example, in linear polyethylene, only a minor fraction of the molecules will be in the most stable zigzag (all-anti) conformation.

These considerations are important in the conformation of proteins (natural polymers). When polypeptides are synthesized—in the laboratory and perhaps also *in vivo*—they are first formed as linear strings (so-called "random coil" conformations, similar to those of a polyethylene). But in polypeptides and proteins additional considerations come into play, notably hydrogen bonding between amino acid residues and hydrophobic forces generated by the reluctance of hydrocarbon side chains to be in contact with the common water solvent. Additional interactions between nonadjacent amino acids may come about because of oxidation of cysteine to cystine residues (2 –SH → –S–S–). These various interactions lead to a folding of the chain into so-called secondary structures, which include a helical (α-helix) conformation (Pauling *et al.*, 1951) stabilized mainly by hydrogen bonding between nonadjacent but close amino acids in the polymer chain, and a doubled-up conformation called a β or pleated sheet stabilized by hydrogen bonding between rather more distant members of the chain folded onto each other. "Tertiary structure" of proteins comprises the combination of α-helices, pleated sheets, and some random-coil areas, which gives rise to their three-dimensional shape.

The distinction between configuration and conformation is usually based on whether the interconversion of the pertinent stereoisomers is slow or fast. Since a barrier of 84 kJ/mole between two species corresponds to an interconversion rate at 25°C of 1.3×10^{-2} sec^{-1}, i.e., a half-life of 1 min, making isolation of the individual species quite difficult, one might say that the division between configuration and conformation comes at barriers of about 84 kJ/mole. However, such a precise distinction is problematic. At lower temperatures, interconversion rates decrease and isomers that differ only in conformation (cf. Fig. 9) may become isolable. Also, the technique of observation matters. Infrared and Raman spectroscopy are "very fast" and thus the vibrational spectra of the conformational isomers of 1,2-dibromoethane are distinct. Nuclear magnetic resonance (NMR) is intermediate, and there are numerous instances where an averaged NMR spectrum is seen at one temperature but spectra for the individual conformers emerge at lower temperatures.

An interesting example of the fluidity of the delineation between configuration and conformation is seen in the biphenyls (Fig. 12). In biphenyl itself, rotation is fast; thus a 3,3′,5,5′-tetrasubstituted biphenyl (Fig. 12A) cannot be resolved into enantiomers, even though conformations in which the two rings are not coplanar are chiral. However, as soon as sizable substituents are introduced at positions 2, 2′, 6, and 6′ (Fig. 12B, X ≠ Y) the compounds become resolvable; they display axial chirality. When X and Y (in Fig. 12B) are different and other than F or CH_3O, the enantiomers are stable. When one of the four substituents is H, however, the compounds are resolvable but usually racemize readily either at room temperature or above by rotation about the Ar–Ar bond. Biphenyls with only two ortho substituents are generally not resolvable unless the substituents are very bulky, as in 1,1′-dinaphthyl (Fig. 12C, Z = H). (The enantiomeric 2,2′-dihydroxybinaphthyls, Fig. 12C, Z = OH, have found manifold uses, e.g., as parts of chiral reagents and chiral catalysts.)

XII. CYCLOALKANES AND THEIR CONFORMATIONS

Before considering conformation in cyclic molecules (which is more complex since rotations about individual

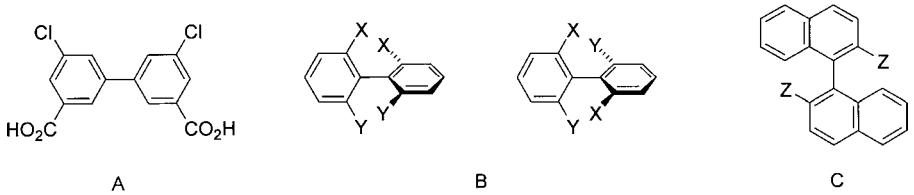


FIGURE 12 Stereoisomerism in biphenyls.

bonds are not independent of each other, unlike in alkane chains) we need to consider the topic of "strain." Already A. von Baeyer in 1885 realized that formation of small rings, such as cyclopropane or cyclobutane, required deformation of the normal tetrahedral or near-tetrahedral bond angle of 109°28′. Thus in cyclopropane the internuclear bond angle is 60°, i.e., it deviates 49°28′ from the normal and this causes angle strain, which in turn destabilizes the cyclopropane molecule. If one takes the contribution of a CH_2 group to the heat of combustion as 658.7 kJ per group [this is the difference in heat of combustion between two large homologous alkanes $CH_3(CH_2)_nCH_3$ and $CH_3(CH_2)_{n+1}CH_3$], the calculated heat of combustion of cyclopropane is $3 \times 658.7 = 1976$ kJ/mole, whereas the experimental value is 2091 kJ/mole; the difference of 115 kJ/mole is a measure of the strain in cyclopropane. Corresponding values are, for cyclobutane, 110 kJ/mole; for cyclopentane, 26.0 kJ/mole; and for cyclohexane, 0.5 kJ/mole. The low value for cyclohexane is at first sight surprising; Baeyer thought that cyclohexane was planar, with bond angles of 120°, and should therefore show strain of $120° - 109°28'$ or 10°32′, though this strain would be due to enlargement rather than diminution of the bond angle. There are two ways of accounting for this discrepancy. First, while strain increases again for the so-called "medium rings" (C_7, 26.2 kJ/mole; C_8, 40.5 kJ/mole; C_9, 52.7 kJ/mole; C_{10}, 51.8 kJ/mole; C_{11}, 47.3 kJ/mole), it diminishes thereafter for the "large rings," e.g., to 8.0 kJ/mole for C_{14}. This is due to the fact that rings other than cyclopropane are actually not planar and there are different sources of strain in these rings (and actually even in cyclopropane). One source is strain due to eclipsing of bonds ("torsional strain"), as explained above for ethane. In planar cyclobutane and cyclopentane, this strain (due to four or five pairs of eclipsed hydrogen atoms, respectively) is large enough to cause these species to be nonplanar, even though this increases angle strain. (A nonplanar polygon has smaller angles than a planar one.) In the larger cycloalkanes, however, where (if they were planar) the angles would be expanded beyond the tetrahedral, puckering actually diminishes not only the torsional or eclipsing strain (see the discussion on ethane above) but also the angle strain. In fact, much of the strain in medium rings is due to nonbonded atoms getting too close to each other; this causes so-called "nonbonded" or van der Waals strain (steric repulsion).

Cyclohexane, which is virtually strain-free, is a special case. In 1890 (only 5 years after Baeyer proposed his strain hypothesis) H. Sachse realized that C_6H_{12} is not planar, but can be constructed from tetrahedral carbon atoms, either in the shape of a chair (Fig. 13A) or that of a boat (Fig. 13B). Today we know that, because of steric repulsion between the hydrogen atoms at C(1) and C(4) pointing inside plus eclipsing strain at C(2, 3) and C(5, 6), the shape in Fig. 13B is actually deformed to a "twist-boat" form (Fig. 13C) and that the chair (Fig. 13A) is the most stable conformer. But it took some 60 years after Sachse for the physical and chemical consequences of the chair shape of cyclohexane to be recognized, by K. Pitzer, O. Hassel, and D. H. R. Barton. Chemically speaking, axial substituents are more hindered (crowded) than equatorial ones and therefore generally less stable, and react more slowly (e.g., in the esterification of acids and alcohol and the hydrolysis of the corresponding esters). Also, the bimolecular elimination reaction (e.g., of H_2O in cyclohexanols or HX in cyclohexyl halides and toluenesulfonates) proceeds more readily when the substituent (OH or X) is axial than when it is equatorial. Barton saw these consequences (and others) of cyclohexane conformation (Eliel *et al.*, 1965) in the rigid cyclohexane systems of steroids and terpenes. Thus in 3-cholestanol (Fig. 14) the equatorial or β isomer is more stable than the axial one (designated α, meaning that the substituent is on the side opposite to the angular methyl groups, whereas β implies that it is on the same side). Also, the β isomer is more easily esterified than the α, but elimination of water to give a cholestene is more facile for the axial α isomer.

In monocyclic cyclohexanes the situation is more complex since the barrier to interconversion of the ring is only

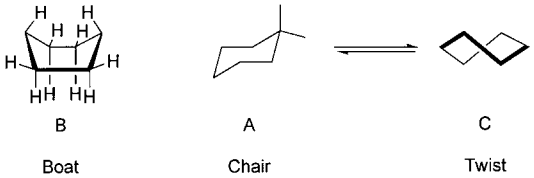


FIGURE 13 Conformations of cyclohexane.

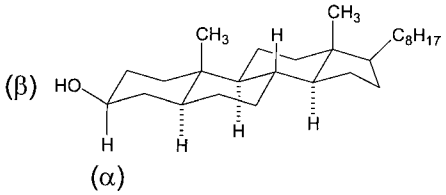


FIGURE 14 Conformation of 3β-cholestanol.

about 42 kJ/mole and interconversion of the two conformers is extremely rapid at room temperature. Thus chlorocyclohexane (Fig. 15, X = Cl) exists in rapid equilibrium between axial and equatorial conformers which differ in free energy by only about 25 kJ/mole, corresponding to 74% of the equatorial and 26% of the axial isomer at 25°C. As expected, in the infrared spectrum there are two C−Cl stretch frequencies, but in the laboratory, chlorocyclohexane, even though a mixture of two conformers, appears as a homogeneous substance with the average properties (such as chemical shifts in NMR) of the two conformers. When one cools the substance to ca. −60°C (the exact temperature required depends on the operational frequency of the NMR instrument), however, two different NMR spectra begin to emerge, and at −150°C the equatorial conformer has actually been crystallized from trideuteriovinyl chloride solution, with concomitant enrichment of the axial isomer in solution.

Equilibria for a large number of monosubstituted cyclohexanes have been determined and tabulated; they were mostly determined by low-temperature ^{13}C NMR spectroscopy (Eliel and Wilen, 1994).

The conformations of piperidine (azacyclohexane) and tetrahydropyran (oxacyclohexane) are qualitatively similar to those of cyclohexane. (Some quantitative differences are seen, for example, in the equatorial preferences of some substituents resulting from dipolar interactions with the ring hetero atom and from the fact that C–N and C–O distances are shorter than C–C in cyclohexane.) These ring systems are important, being found in alkaloids and hexose sugars, respectively.

Because of torsional (eclipsing) strain, cyclobutane and cyclopentane are not planar. Cyclobutane is wing-shaped; cyclopentane oscillates among a number of low-energy conformations of which the envelope (four carbon atoms in a plane, one out of plane) and the half-chair (three adjacent carbons in a plane, the fourth above that plane and the fifth below) are the most symmetrical. The barrier between these conformations is very low; thus their rapid interconversion, which involves up-and-down motions of successive adjacent carbon atoms, has the appearance of a bulge moving around the rings; this process has therefore been named "pseudorotation." Higher cycloalkanes from C(7) on display families of conformations which are separated by barriers similar to those in cyclohexane, but within a given family there may be several individual members interconverted by pseudorotation. This subject is discussed in detail in Eliel and Wilen (1994).

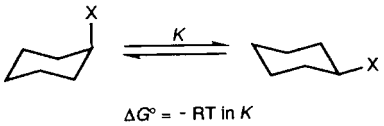


FIGURE 15 Conformational inversion of substituted cyclohexane. [Reprinted with permission from Eliel, E. L., and Wilen, S. H. (1994). "Stereochemistry of Organic Compounds," Wiley, New York.]

XIII. CHIROPTICAL PROPERTIES. ENANTIOMERIC PURITY

By "chiroptical properties" are meant optical properties that differ between enantiomers and can be used to characterize them. They comprise optical rotation, optical rotatory dispersion (ORD), and circular dichroism (CD).

Optical rotation has already been discussed. Because of its critical dependence on solvent (including cosolvents, such as ethanol in chloroform), temperature, concentration, and the potential presence of impurities, especially chiral impurities, in the sample, experimental determination of [α] requires considerable care and many of the values given in the literature cannot be trusted. This is unfortunate since it is often desirable to determine the "enantiomeric purity" of a sample to see whether the desired enantiomer is obtained free of the other. (For example, in pharmaceutical chemistry one wants to obtain the pure eutomer free of the distomer; see above.) Since in all but a few cases optical rotation is proportional to the fraction of the major enantiomer in the total substance, one might expect enantiomeric or optical purity to be equal to $100[\alpha_{obs}]/[\alpha_{max}]$%, where $[\alpha_{max}]$ is the (presumed known) specific rotation of an enantiomerically pure sample. However, this is true only if both values have been accurately determined under exactly the same conditions (solvent, temperature, etc.).

Because of this difficulty, other, more reliable methods of determining enantiomeric purity (now no longer called "optical purity") have been developed. Basically these depend on converting the enantiomers into diastereomers, either by covalent chemical bonding or by complexing in some fashion, with another, usually enantiomerically pure chiral auxiliary. (For some methods the auxiliary need not even be enantiomerically pure.) Once the enantiomers have been converted into diastereomers, their ratio can be determined by NMR or chromatographic methods,

including NMR in a chiral solvent or with a chiral complexing agent, or by chromatography of unmodified enantiomers on a chiral stationary phase. The methods just described provide the mole fractions of each individual enantiomer, which may be expressed as an enantiomer ratio (n_+/n_- or n_-/n_+). However, because optical purity had been used in the past, it is more common to use an expression for "enantiomeric excess" (e.e.): e.e. $= |(n_+ - n_-)|/(n_+ + n_-)$. Since this represents the mole fraction of the major enantiomer diminished by that of the minor one, it is equal to the above-defined optical purity.

Although there is no direct relation between sign of rotation and configuration, it may now become possible to infer configuration from optical rotation data (Kondra, Wipf, and Beratan, 1998). The method is based on van't Hoff's "optical superposition" rule, which says that in a molecule containing several chiral centers, the total molar rotation ($\Phi = 100\alpha/MW$) is the sum of the contributions of each chiral center. As originally proposed the rule had several shortcomings, including (1) that it would require model compounds of known absolute configuration to determine the contribution of a given chiral center and (2) it does not hold when the centers are close to each other and thus influence each other's contributions. This second limitation means that for closely connected chiral centers, it is necessary to determine the contribution of a segment containing all of these centers. The first problem is being attacked by performing *a priori* computations of the molar rotation contribution of individual chiral centers or appropriate groupings thereof (Kondra, Wipf, and Beratan, 1998). This is becoming possible due to the advances in quantum chemical calculations (e.g., by density functional methods) and the increasing power of computers to handle computationally demanding problems.

Other chiroptical techniques to infer configuration are optical rotatory dispersion (ORD) and circular dichroism (CD). ORD relates to the change in optical rotation with the wavelength of the light employed in the measurement. Normally the absolute value of rotation increases as wavelength becomes shorter; observation at shorter wavelengths is thus a convenient way to increase rotation (and thereby the accuracy of measuring it) when α_D is small. However, as the wavelength approaches that of a UV absorption band (e.g., of C═O in a ketone), its absolute value (whether positive or negative) suddenly drops precipitously, passes through zero near the UV absorption maximum, reverses sign rapidly approaching another extremum (of opposite sign to the first), and then gradually declines. This phenomenon of rapid change at the UV maximum is called the "Cotton effect" after the French scientist who discovered it. Similarity in Cotton effects of related compounds, one of known and one of unknown configuration, can sometimes be used to assign the configuration of the unknown. However the use of ORD has largely been superseded by the simpler to interpret CD (Nakanishi *et al.*, 2000). While it might appear that plane-polarized light is achiral, it may actually be considered to be a superposition of right- and left-circularly polarized light in which the sense of polarization changes as a right-handed or left-handed helix along the direction of propagation of the light beam. If the right- and left-circularly polarized beams proceed at the same speed, the result is light polarized in an unchanging plane, but if one of the two generating beams moves faster than the other, the plane of polarization will keep turning as the light propagates, i.e., there will be optical rotation. Since the speed of light depends on the refractive index of the medium it traverses, the polarization is thus caused by unequal refractive indices for right- and left-circularly polarized light. There are devices that can produce right- and left-circularly polarized light beams separate from each other. Using such beams, it is found that not only the refractive indices, but also the absorption coefficients for the two beams differ. The phenomenon resulting from this difference in absorption is called "circular dichroism" and manifests itself in what looks similar to an absorption curve in the UV, except that it is signed. (In fact its maximum or minimum occurs at the wavelength of the UV maximum.) Comparison of CD absorption spectra can be used to infer configuration similarly as was the case for ORD; however, CD spectra are better resolved than ORD spectra since there is less band overlap resulting from multiple UV absorption maxima.

CD is also very useful in throwing light on conformation. Thus the (weak) CD absorption spectrum of a random-coil polypeptide chain is essentially a superposition of the spectra of the individual constituting amino acids. However, when secondary structure comes into play, as in an α helix of β-pleated sheet, a large and characteristic increase in CD absorption is observed, which, in turn, allows one to infer the nature of the secondary structure, if any. CD is used to infer not only protein but also nucleic acid and polysaccharide conformation (Fasman, 1996).

XIV. PROCHIRALITY

The phenomenon of prochirality (Mislow and Raban, 1967) is important both in NMR spectroscopy and in enzyme chemistry. An atomic center (e.g., a tetrahedral carbon atom) in a molecule is considered "prochiral" if replacing one of two identical ligands at the center by a different one not previously attached to that center produces a chiral center. Thus the carbon atom in CH_2ClBr is prochiral since hypothetical replacement of one of the hydrogen atoms by deuterium yields CHDClBr, which is chiral. The apparently identical (or "homomorphous," from Greek "homos," same, and "morphe," form) hydrogen atoms in CH_2ClBr are in fact distinct; they are

called "heterotopic" (from Greek "heteros," different, and "topos," place). In contrast, the carbon atom in CH_3Cl is not prochiral since replacement of H by, say, D would produce CH_2DCl, which remains achiral.

In the former case (CH_2ClBr) replacement of one or other of the two hydrogen atoms gives rise to enantiomeric products. The hydrogen atoms are therefore called "enantiotopic." In the latter case, replacement gives the same compound and the hydrogens in CH_3Cl are called "homotopic." In a molecule such as $CH_2BrCHOHCO_2H$, replacement of one of the terminal hydrogens by, say, chlorine would give one or other of the diastereomers of $CHBrClCHOHCO_2H$; in this case the terminal hydrogens are said to be diastereotopic. These definitions of homotopic, enantiotopic, and diastereotopic ligands also provide a means for their recognition: Replacement of one of two or more homotopic ligands by a different ligand gives identical products, analogous replacement of enantiotopic ligands gives enantiomeric products, and such replacement of diastereotopic ligands gives diastereomeric products. There is also a symmetry criterion which may be applied to the appropriate molecules above: Homotopic ligands in a molecule are interchanged by operation of both simple symmetry axes and symmetry planes; enantiotopic ligands are interchanged by operation of a symmetry plane but not by operation of a simple symmetry axis, and diastereotopic ligands are interchanged neither by symmetry axes nor by symmetry planes.

Diastereotopic ligands (e.g., protons or C-13 atoms) generally display distinct signals in NMR spectra, but homotopic and enantiotopic ligands have coincident (identical) signals (except possibly in the case of enantiotopic ligands, in a chiral solvent, or in the presence of a chiral complexing agent) since NMR is an achiral technique. Both enantiotopic and diastereotopic ligands may be distinguished by enzymes (which are chiral). Thus in citric acid, $HO_2CCH_2C(OH)(CO_2H)CH_2CO_2H$, all four methylene hydrogen atoms are distinguished by enzymes in the citric acid cycle (to demonstrate this distinction, they must be individually labeled as deuterium atoms). On the other hand, the CH_2 groups are pairwise identical in NMR (e.g., C-13) but the geminal hydrogen atoms in each are diastereotopic and provide an $(AB)_2$ system in the proton NMR spectrum. Further details may be found in Eliel (1982) and Eliel and Wilen (1994).

SEE ALSO THE FOLLOWING ARTICLES

BIOPOLYMERS • ENZYME MECHANISMS • NUCLEAR MAGNETIC RESONANCE • ORGANIC CHEMICAL SYSTEMS, THEORY • ORGANIC CHEMISTRY, SYNTHESIS • PHYSICAL ORGANIC CHEMISTRY • PROTEIN STRUCTURE • PROTEIN SYNTHESIS • RHEOLOGY OF POLYMERIC LIQUIDS

BIBLIOGRAPHY

Addadi, L., Berkovitch-Yellin, Z., Weissbuch, I., Lahav, M., and Leiserowitz, L. (1986). "A link between macroscopic phenomena and molecular chirality: Crystals as probes for the direct assignment of absolute configuration of chiral molecules. *In* "Topics in Stereochemistry," Vol. 16, pp. 1–85, Wiley, New York.

Cahn, R. S., Ingold, C., and Prelog, V. (1966). "Specification of molecular chirality," *Angew. Chem. Int. Ed. Engl.* **5,** 385–415 .

Eliel, E. L. (1982). "Prostereoisomerism (prochirality)."*In* "Topics in Current Chemistry," Vol. 105, pp. 1–76, Springer-Verlag, Heidelberg.

Eliel, E. L., and Wilen, S. H. (1994). "Stereochemistry of Organic Compounds," Wiley, New York.

Eliel, E. L., Allinger, N. L., Angyal, S. J., and Morrison, G. A. (1965). "Conformational Analysis," Wiley, New York [reprinted (1981), American Chemical Society, Washington, DC.]

Fasman, G. D. (ed.). (1996). "Circular Dichroism and the Conformational Analysis of Biomolecules," Plenum Press, New York.

Gawley, R. E., and Aubé, J. (1996). "Principles of Asymmetric Synthesis," Pergamon Press, Oxford.

Hegstrom, R. A., and Kondepudi, D. K. (1990). "The handedness of the universe," *Sci. Am.* **262**(January), 108–115.

Jacques, J., Collet, A., and Wilen, S. H. (1981). "Enantiomers, Racemates and Resolutions," Wiley, New York.

Juaristi, E. (ed.). (1995). "Conformational Behavior of Six-Membered Rings," VCH, New York.

Kagan, H. B., and Fiaud, J. C. (1988). "Kinetic resolution." *In* "Topics in Stereochemistry," Vol. 18, pp. 249–330, Wiley, New York.

Klyne, W., and Prelog, V. (1960). "Description of stereochemical relationships across single bonds," *Experientia* **16,** 521–523.

Kondru, R. K., Wipf, P., and Beratan, D. N. (1998). "Atomic contributions to the optical rotation angle as a quantitative probe of molecular chirality," *Science* **282,** 2247–2250; *id*. (1998). Theory-assisted determination of absolute stereochemistry for complex natural products vid computation of molecular rotation angle, *J. Am. Chem. Soc.* **120,** 2204–2205.

Mislow, K., and Raban, M. (1967). "Stereoisomeric relationships of groups in molecules." *In* "Topics in Stereochemistry," Vol. 1, pp. 1–38, Wiley, New York.

Nakanishi, K., Berova, N., and Woody, R. W. (eds.). (2000). "Circular Dichroism: Principles and applications," Wiley-VCH, New York.

Pauling, L., Corey, R. B., and Branson, H. R. (1951). "The structure of proteins: Two hydrogen bonded helical configurations of the polypeptide chain," *Proc. Natl. Acad. Sci. U.S.A.* **37,** 205–211.

Prelog, V., and Helmchen, G. (1982). "Basic principles of the CIP system and proposals for a revision," *Angew. Chem. Int. Ed. Engl.* **21,** 567–583.

Ramsay, O. B. (1981). "Stereochemistry," Heyden & Son, Philadelphia.

Sih, C. J., and Wu, S.-H. (1989). "Resolution of enantiomers via biocatalysis." *In* "Topics in Stereochemistry," Vol. 19, pp. 63–125, Wiley, New York.